AN SU(2) ANOMALY

Edward WITTEN [1]
Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544, USA

Received 14 July 1982

A new restriction on fermion quantum numbers in gauge theories is derived. For instance, it is shown that an SU(2) gauge theory with an odd number of left-handed fermion doublets (and no other representations) is mathematically inconsistent.

It has been a long-standing puzzle to elucidate the properties of an SU(2) gauge theory with a single doublet of left-handed (Weyl) fermions. This theory defies simple phenomenological descriptions. There is no obvious attractive channel in which a fermion condensate could form, consistent with Fermi statistics and Lorentz invariance. But it is hard to believe that the fermions could remain massless in the presence of strong SU(2) gauge forces at long distances.

This puzzle persists (in the absence of other representations) whenever the number of elementary fermion doublets is odd. An even number of doublets, even if they have zero bare mass, could pair up and become massive Dirac fermions through spontaneous chiral symmetry breaking. With an odd number of elementary doublets, however, there would always be one massless doublet left over after any assumed chiral symmetry breaking, as long as the SU(2) gauge symmetry remains unbroken.

Of course, there is no real paradox here. Perhaps our heuristic pictures of strongly interacting gauge theories are inadequate. However, the facts noted above do suggest that something is strange about an SU(2) gauge theory with an odd number of elementary doublets. The purpose of this paper is to determine precisely what is strange about these theories; we will see that they are mathematically inconsistent! The inconsistency arises from a problem somewhat analogous to the Adler–Bell–Jackiw anomaly.

Although a hamiltonian approach exists, let us first look at this problem from the point of view of euclidean functional integrals. The starting point is the fact [1] that the fourth homotopy group of SU(2) is nontrivial,

$$\pi^4(\mathrm{SU}(2)) = \mathrm{Z}_2. \tag{1}$$

[Note that we are dealing with the *fourth* homotopy group, while the *third* homotopy group, $\pi^3(\mathrm{SU}(2)) = \mathrm{Z}$, has entered in instanton studies [2]. The analogue of π^4 has entered in some recent studies [3] of 2 + 1 dimensional models.] Eq. (1) means that in four-dimensional euclidean space, there is a gauge transformation $U(x)$ such that $U(x) \to 1$ as $|x| \to \infty$, and $U(x)$ "wraps" around the gauge group in such a way that it cannot be continuously deformed to the identity. The fact that the homotopy group is Z_2 means that a gauge transformation that wraps twice around SU(2) in this way can be deformed to the identity. We will not need the detailed form of $U(x)$.

The existence of the topologically non-trivial mapping $U = U(x)$ means that when we carry out the euclidean path integral

$$\int (\mathrm{d}A_\mu) \exp\left(-\frac{1}{2g^2} \int \mathrm{d}^4x \, \mathrm{tr}\, F_{\mu\nu}F^{\mu\nu}\right), \tag{2}$$

we are actually double counting. For every gauge field A_μ, there is a conjugate gauge field

$$A_\mu^{\mathrm{U}} = U^{-1}A_\mu U - \mathrm{i}U^{-1}\partial_\mu U,$$

which makes exactly the same contribution to the functional integral. There is no way to eliminate this

[1] Supported in part by the National Science Foundation under Grant No. PHY80-19754.

double counting because A_μ and A_μ^U lie in the same sector of field space; A_μ^U can be reached continuously from A_μ without passing through singularities or infinite action barriers. But, in the absence of fermions, the double counting is harmless and cancels out when one calculates vacuum expectation values.

Now let us include fermions. Introducing, say, a single doublet of left-handed fermions, we now must deal with

$$Z = \int d\psi\, d\bar{\psi} \int dA_\mu$$

$$\times \exp\left(-\int d^4x \left[(1/2g^2)\, \mathrm{tr}\, F_{\mu\nu}^2 + \bar{\psi} i \not{D} \psi\right]\right) . \qquad (3)$$

We would like to integrate out the fermions and discuss the effective theory with the fermions eliminated.

As is well known, for a theory with a doublet of *Dirac* fermions, the basic integral is

$$\int (d\psi\, d\bar{\psi})_{\mathrm{Dirac}} \exp(\bar{\psi} i \not{D} \psi) = \det i\not{D}. \qquad (4)$$

Here the right-hand side is, formally, the infinite product of all eigenvalues of the hermitian operator $i\not{D} = i\gamma_\mu D^\mu$. Certain theories – those that are afflicted with Adler–Bell–Jackiw anomalies – are ill-defined because it is impossible to renormalize this formal product so as to get a gauge invariant answer. However, a doublet of Dirac fermions could have a gauge invariant bare mass; this means that Pauli–Villars regularization is available, and hence that the determinant in (4) can be defined satisfactorily. This determinant is completely gauge invariant – invariant both under infinitessimal gauge transformations and under the topologically non-trivial gauge transformation U discussed earlier.

Now, with the gauge group SU(2), a doublet of Dirac fermions is exactly the same as two left-handed or Weyl doublets. Hence the fermion integration with a single Weyl doublet would give precisely the square root of (4):

$$\int (d\psi\, d\bar{\psi})_{\mathrm{Weyl}} \exp(\bar{\psi} i \not{D} \psi) = (\det i\not{D})^{1/2}. \qquad (5)$$

But an ambiguity arises here; the square root has two signs. As we will see, this ambiguity leads to trouble.

Picking a particular gauge field A_μ, we are free to define in an arbitrary way the sign of $(\det i\not{D})^{1/2}$ for this field. Once this is done, there is no further freedom; to satisfy the Schwinger–Dyson equations we must define the fermion integral $(\det i\not{D})^{1/2}$ to vary smoothly as A_μ is varied.

Defined in this way $(\det i\not{D})^{1/2}$ is certainly invariant under infinitessimal gauge transformations – since the sign does not change abruptly. But nothing guarantees that $(\det i\not{D})^{1/2}$ is invariant under the topologically non-trivial gauge transformation U. In fact, as we will see, $(\det i\not{D})^{1/2}$ is odd under U. We will see that for any gauge field A_μ,

$$[\det i\not{D}(A_\mu)]^{1/2} = -[\det i\not{D}(A_\mu^U)]^{1/2}. \qquad (6)$$

In other words, if one continuously varies the gauge field from A_μ to A_μ^U, one ends up with the opposite sign of the square root.

Before explaining why eq. (6) is valid, let us first discuss why it results in the mathematical inconsistency of the SU(2) theory with a single left-handed doublet. The partition function would be

$$Z = \int dA_\mu (\det i\not{D})^{1/2} \exp\left(-\frac{1}{2g^2}\int d^4x\, \mathrm{tr}\, F_{\mu\nu}^2\right). \quad (7)$$

But this vanishes identically, because the contribution of any gauge field A_μ is exactly cancelled by the equal and opposite contribution of A_μ^U! Likewise the path integral Z_X with insertion of any gauge invariant operator X is identically zero. So expectation values are indeterminate, $\langle X \rangle = Z_X/Z = 0/0$. For this reason, the theory is ill-defined.

One cannot avoid this problem by taking the absolute value of $(\det i\not{D})^{1/2}$; the resulting theory would not obey the Schwinger–Dyson equations. Nor can one consistently integrate over only "half" of field space, since A_μ and A_μ^U are continuously connected.

It remains to explain eq. (6). For convenience, take space–time to be a sphere of large volume so that the spectrum of $i\not{D}$ is discrete. We may as well assume there are no zero eigenvalues since otherwise $\det i\not{D}(A_\mu)$ vanishes and (6) is certainly true. The eigenvalues of $i\not{D}$ are real and (fig. 1) for every eigenvalue λ there is an eigenvalue $-\lambda$, since if $i\not{D}\psi = \lambda\psi$, then $i\not{D}(\gamma_5\psi) = -\lambda(\gamma_5\psi)$.

Taking the square root of $\det i\not{D}$ means that we want the product of only half of the eigenvalues, not all of them. We may suppose that for every pair of eigenvalues $(\lambda, -\lambda)$ we pick one or the other, but not both. For instance, for a particular gauge field A_μ we

(1)

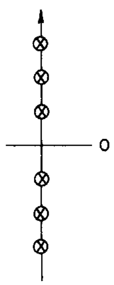

Fig. 1. The spectrum of the Dirac operator for a particular gauge field A_μ. The square root of the determinant may be defined – for this particular gauge field – as the product of the positive eigenvalues.

may define $(\det i\not{D})^{1/2}$ to be the product of the positive eigenvalues (fig. 1).

Now imagine varying the gauge field along a continuous path in field space from A_μ to A_μ^{U}. For instance, one may consider the gauge field $A_\mu^{t} = (1-t)A_\mu + tA_\mu^{\mathrm{U}}$, with t varied smoothly from zero to one. Let us follow the flow of the eigenvalues as a function of t. The spectrum of $i\not{D}$ is precisely the same at $t = 1$ as it is at $t = 0$. However, the individual eigenvalues may rearrange themselves as t is varied from zero to one.

As will be explained, the Atiyah–Singer index theorem predicts that such a rearrangement occurs. The simplest rearrangement allowed by the theorem is indicated in fig. 2. A single pair of eigenvalues $\{\lambda(t), -\lambda(t)\}$ cross at zero and change places as t is varied form zero to one.

In particular, one of the eigenvalues which was positive at $t = 0$ is negative by the time $t = 1$. If at $t = 0$, $(\det i\not{D})^{1/2}$ was defined as the product of the positive eigenvalues, then, following the eigenvalues continuously, by the time we reach $t = 1$ $(\det i\not{D})^{1/2}$ is the product of many positive eigenvalues and a single negative one. This means that $(\det i\not{D})^{1/2}$ has the opposite sign at $t = 1$ from its value at $t = 0$. The square root vanishes when the eigenvalue pair passes through zero ($t = t_0$ in fig. 2) and is negative for $t > t_0$.

(2)

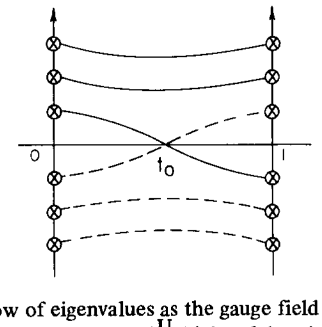

Fig. 2. The flow of eigenvalues as the gauge field is varied from A_μ (left of drawing) to A_μ^{U} (right of drawing). The square root of the determinant may be defined as the product of the eigenvalues indicated by solid lines; it vanishes and changes sign at $t = t_0$.

The Atiyah–Singer theorem permits more complicated rearrangements of eigenvalues, but the number of positive eigenvalues that become negative as t is varied from 0 to 1 is always *odd.* This is the basis for eq. (6).

The connection between the index theorem and the flow of eigenvalues is well known in mathematics [4] and has been discussed in the physics literature [5]. What is relevant for our problem is a slightly exotic form of the index theorem, namely the mod two index theorem for a certain *five*-dimensional Dirac operator [6].

Consider the five-dimensional Dirac equation for an SU(2) doublet of fermions,

$$\not{D}^{(5)}\Psi = \sum_{i=1}^{5} \gamma^i \left(\partial_i + \sum_{a=1}^{3} A_i^a T^a\right) \Psi = 0. \tag{8}$$

The spinor Ψ has eight components because the spinor representation of O(5) is four dimensional while an SU(2) doublet has two components.

The spinor representation of O(5) is pseudo-real, rather than real, and the doublet of SU(2) is likewise pseudo-real. But the tensor product of the spinor representation of O(5) with the doublet of SU(2) is a *real* representation of O(5) × SU(2). This means that in (8), we can take the gamma matrices γ^i to be real, symmetric 8 × 8 matrices while the anti-hermitian generators T^a of SU(2) are real, anti-symmetric matrices [‡1].

The five-dimensional Dirac operator $\not{D}^5$ for an SU(2) doublet is therefore a real, antisymmetric opera-

‡1 In fact, one can arrange Ψ as a two-component column vector of quaternions $\Psi = \binom{\psi_1}{\psi_2}$ – which would have eight real components. Such a column vector can be acted on from the left by a 2 × 2 unitary matrix of quaternions [making the group Sp(2) or O(5)] and on the right by a unitary quaternion [the group Sp(1) or SU(2)]. This is the desired eight-dimensional real representation of O(5) × SU(2).

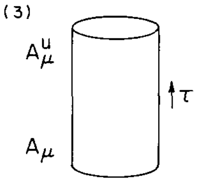

Fig. 3. A five-dimensional cylinder, $S^4 \times R$, on which an instanton-like gauge field is defined.

tor, acting on an infinite dimensional space. The eigenvalues of such a real, antisymmetric operator either vanish or are imaginary and occur in complex conjugate pairs. When the gauge field A_μ is varied, the number of zero eigenvalues of $D\!\!\!/^5$ can change only if a complex conjugate pair of eigenvalues moves to – or away from – the origin. The number of zero eigenvalues of $D\!\!\!/^5$ modulo two is therefore a topological invariant. It is known as the mod two index of the Dirac operator.

Now, consider a five dimensional cylinder $S^4 \times R$ (fig. 3). Let x^μ, $\mu = 1, ..., 4$, be coordinates for S^4 while the position in the "time" direction (position in R) is called τ. Consider the following five-dimensional instanton-like SU(2) gauge field. For all x^μ and τ, $A_\tau = 0$. But $A_\mu(X^\sigma, \tau)$, $\mu = 1, ..., 4$, is – for each τ – a four-dimensional gauge field described as follows. For $\tau \to -\infty$ $A_\mu(x^\sigma, \tau)$ approaches the four-dimensional gauge field A_μ of our previous discussion ($t = 0$ in fig. 2). For $\tau \to +\infty$ $A_\mu(x^\sigma, \tau)$ approaches what we previously called A^U_μ ($t = 1$ in fig. 2). As τ varies from $-\infty$ to $+\infty$, $A_\mu(X^\sigma, \tau)$ varies *adiabatically* from A_μ to A^U_μ, along the same path in field space considered in fig. 2.

The mod two Atiyah–Singer index theorem predicts that the number of zero modes in this five-dimensional gauge field is odd – equal to one modulo two [‡2].

On the other hand, the number of zero modes of $D\!\!\!/^5$, modulo two, can be calculated in terms of the eigenvalue flow of fig. 2. The Dirac equation $D\!\!\!/^5 \Psi = 0$ can be written

$$d\Psi/d\tau = -\gamma^\tau D\!\!\!/^4 \Psi, \tag{9}$$

where $D\!\!\!/^4 = \Sigma_{i=1}^4 \gamma^\mu D_\mu$ is – at each τ – a four-dimensional Dirac operator.

Since $A_\mu(x^\sigma, \tau)$ evolves adiabatically in τ, (9) can be solved in the adiabatic approximation. We write $\Psi(x^\mu, \tau) = F(\tau)\, \phi^\tau(x^\mu)$, where $\phi^\tau(x^\mu)$ is a smoothly evolving solution of the eigenvalue equation

$$\gamma^\tau D\!\!\!/^4 \phi^\tau(x^\mu) = \lambda(\tau)\phi^\tau(x^\mu). \tag{10}$$

The eigenvalues $\lambda(\tau)$ evolve on the curves of fig. 2 ($iD\!\!\!/^4$ and $\gamma^\tau D\!\!\!/^4$ have the same spectrum). In the adiabatic limit, eq. (9) now reduces to $dF/d\tau = -\lambda(\tau)F(\tau)$, and the solution is

$$F(\tau) = F(0) \exp\left(-\int_0^\tau d\tau' \, \lambda(\tau')\right). \tag{11}$$

This is normalizable only if λ is positive for $\tau \to +\infty$, and negative for $\tau \to -\infty$.

In the adiabatic approximation, the number of zero eigenvalues of $D\!\!\!/^5$ is therefore equal to the number of eigenvalue curves in fig. 2 which pass from negative to positive values (or from positive to negative values) between $t = 0$ and $t = 1$. When corrections to the adiabatic approximation are considered, this gives the number of zero eigenvalues modulo two.

The Atiyah–Singer theorem, which requires that $D\!\!\!/^5$ has an odd number of zero eigenvalues, therefore implies that the number of eigenvalue curves that pass from positive to negative values in fig. 2 is odd. This is precisely what we needed to show that $(\det iD\!\!\!/)^{1/2}$ is odd under the topologically non-trivial gauge transformation U.

Now let us consider some generalizations. With n left-handed fermion doublets, the fermion integration would give $(\det iD\!\!\!/)^{n/2}$. If n is even, the sign of the square root does not matter, but if n is odd, the theory suffers from the same inconsistency as before.

This persists even if additional gauge or Yukawa couplings are added to an SU(2) gauge theory. Since the fermion integration is necessarily either even or odd under U, if it is odd in a pure SU(2) gauge theory, it remains odd if additional gauge or Yukawa couplings are smoothly switched on. In particular, the standard $SU(3) \times SU(2) \times U(1)$ model of strong, weak, and electromagnetic interaction would be inconsistent if the number of left-handed fermion doublets were odd.

If one considers theories with SU(2) representations of isospin bigger than one half, the Atiyah–Singer

‡2 Actually, in a special case one can easily find the zero mode. If one conformally compactifies $S^4 \times R$ to the five-sphere S^5, then on S^5 one can choose the instanton field to be invariant (up to a gauge transformation) under an SU(3) subgroup of the symmetry group O(6) of the five-sphere. The fermion zero mode is then the unique SU(3) invariant spinor field that can be defined.

theorem gives the following result. If one normalizes the SU(2) generators conventionally so that tr $T_3^2 = 1/2$ in the doublet representation, then the fermion integration is even or odd depending on whether tr T_3^2, evaluated among all the left-handed fermions, is an integer or half-integer. The inconsistent theories are those where tr T_3^2 is a half-integer. (In an ordinary instanton field, the number of fermion zero modes is 2 tr T_3^2, so the inconsistent theories are precisely those with an odd number of fermion zero modes in an instanton field.)

Considering gauge groups other than SU(2), we have

$$\pi^4(\mathrm{SU}(N)) = 0, \quad N > 2,$$

$$\pi^4(\mathrm{O}(N)) = 0, \quad N > 5,$$

$$\pi^4(\mathrm{Sp}(N)) = \mathrm{Z}_2, \quad \text{any } N. \tag{12}$$

Thus non-trivial conditions arise only for Sp(N) groups.

Finally, let us note how this appears in a hamiltonian framework. Space permits only a brief statement of results.

From a hamiltonian viewpoint, one introduces the group G consisting of all gauge transformations $U(x, y, z)$ defined in *three*-dimensional space such that $U(\boldsymbol{x}) \to 1$ as $|\boldsymbol{x}| \to \infty$.

The fact that $\pi^4(\mathrm{SU}(2)) = \mathrm{Z}_2$ means that $\pi^1(\mathrm{G}) = \mathrm{Z}_2$. For the topologically non-trivial gauge transformation $U = U(x, y, z, t)$ that we have discussed is – at each t – an element of G. At $t \to -\infty$ or $t \to +\infty$ it is the identity in G; varying t from $-\infty$ to $+\infty$, U describes a loop in G which cannot be deformed away.

In canonical quantization, one encounters operators

$$Q^a(x) = g^{-2} D_i F^a_{0i}(x) - \bar{\psi}\gamma^0 T^a \psi, \tag{13}$$

which are generators of the Lie algebra of G. However, when a group – in this case G – is not simply connected, a representation of the Lie algebra does not necessarily provide a representation of the group. In general one gets a representation only of the simply connected covering group $\bar{\mathrm{G}}$. Since $\pi^1(\mathrm{G}) = \mathrm{Z}_2$, the center of $\bar{\mathrm{G}}$ has a single non-trivial element P.

In quantum field theory, P – being in the center of $\bar{\mathrm{G}}$ – commutes with all fields and therefore is a c-number. Since $P^2 = 1$, we must have $P = +1$ or $P = -1$ (as an operator statement) in any given field theory. The theories in which the fermion integration is odd under U are the theories in which $P = -1$.

Theories with $P = -1$ are inconsistent for the following reason. According to Gauss's law, physical states $|\psi\rangle$ must be gauge invariant, obeying $Q^a|\psi\rangle = 0$ and hence $P|\psi\rangle = |\psi\rangle$. If $P = -1$, there are no states in the entire Hilbert space that obey Gauss's law.

Similar behavior can be seen in the models of ref. [3] by means of canonical quantization. This was one motivation for the present work.

I would like to thank S. Coleman and H. Georgi for discussions of the SU(2) theory with an odd number of doublets, I. Affleck and J. Harvey for discussions of fermion integration, and R. Jackiw for discussions about the models of ref. [3]. I also wish to thank W. Browder and J. Milnor for discussions about topology.

References

[1] S.T. Hu, Homotopy theory (Academic Press, New York, 1959) Ch. 11;
H. Toda, Composition methods in homotopy groups of spheres (Princeton U.P., Princeton, 1962).
[2] A.A. Belavin, A.M. Polyakov, A. Schwarz and Y. Tyupkin, Phys. Lett. 59B (1975) 85;
C. Callan R. Dashen and D.J. Gross, Phys. Lett. 63B (1976) 334;
R. Jackiw and C. Rebbi, Phys. Rev. Lett. 37 (1976) 334.
[3] S. Deser, R. Jackiw and S. Templeton, MIT preprint (1982).
[4] M.F. Atiyah, V. Patodi and I. Singer, Math. Proc. Camb. Philos. Soc. 79 (1976) 71.
[5] C.G. Callan, R. Dashen and D.J. Gross, Phys. Rev. D17 (1978) 2717;
J. Kiskis, Phys. Rev. D18 (1978) 3061.
[6] M.F. Atiyah and I.M. Singer, Ann. of Math. 93 (1971) 119.