Gauge Anomalies in an Effective Field Theory[⋆]

John Preskill

California Institute of Technology, Pasadena, CA 91125

Abstract

A four–dimensional gauge theory with anomalous fermion content can be consistently quantized, provided that at least some gauge fields are permitted to have nonvanishing masses. Such a theory is nonrenormalizable; there is a maximal value of the ultraviolet cutoff Λ, beyond which the locality of the theory breaks down. The maximal Λ can be estimated in perturbation theory, and has a qualitatively different character in Abelian and non–Abelian anomalous gauge theories.

1. Introduction

When field theory or string theory is used to describe the interactions of elementary particles, it is often convenient to "integrate out" heavy degrees of freedom and construct an *effective field theory* that describes the interactions of the remaining light excitations at sufficiently low energy. Indeed, whenever we apply field theory to particle phenomenology, we are making use of an effective field theory — we do not and need not assume that our field theory accurately describes physics at arbitrarily short distances. Rather, we expect that an accurate description of very–short–distance physics can be achieved only by introducing new degrees of freedom and new interactions. We should always imagine, therefore, that the field theory that we are working with comes equipped with a *cutoff* Λ; Λ is the mass scale beyond which our effective field theory is no longer expected to apply.

Even though it is not intended to provide a complete description of physics at arbitrarily short distances, an effective field theory must satisfy some nontrivial consistency requirements. Often invoked[1] is the requirement that chiral gauge anomalies must cancel. This requirement restricts the representation content under the gauge group of the fermions in the theory. Our objective in this paper is to determine the precise conditions under which anomalies must cancel in order for an effective gauge field theory to make sense.

Why is anomaly cancellation required for consistency? Two reasons are often cited. The first reason is that anomalies cause a loss of *unitarity* or *Lorentz invariance*. The point is that the gauge anomaly is a breakdown of gauge invariance at the quantum level. But we need gauge invariance to establish the equivalence of the covariant gauge and physical gauge formulations of a gauge theory, and thus to assure that the theory can be so formulated as to satisfy both unitarity and Lorentz invariance simultaneously[2]. The second reason is that anomalies cause a loss of *renormalizability*. The gauge anomaly causes the divergence structure of the theory, softened by gauge invariance, to become more severe, so that an infinite number of counterterms are generated.

1

From the effective field theory viewpoint, it is clearly important to distinguish between these two problems. A loss of unitarity — the failure of negative metric states to decouple in covariant gauges — renders the effective field theory itself inconsistent. But a loss of renormalizability signifies that, while an effective field theory with cutoff Λ makes sense when Λ is small enough, it is not possible to push Λ up arbitrarily high without our theory suffering a loss of *calculability*; at some mass scale, new physics must be encountered.

In fact, I would argue that nonrenormalizability is the *only* genuine shortcoming of an effective theory with uncanceled gauge anomalies. Unitarity and Lorentz invariance are easy to achieve, if we are willing to allow the gauge fields to acquire masses. To see this, we need only consider a model gauge theory with a set of fermions in a reducible anomaly–free representation, and in which the gauge symmetry is spontaneously broken. If the spontaneous breakdown of the gauge symmetry is driven by a Higgs field that acquires a vacuum expectation value at the mass scale v, then we may consider the effective field theory that describes the physics of this model at energies well below v. At least some of the gauge fields acquire masses $\mu \sim ev$, which may be quite small compared to v if the gauge coupling e is weak. If the Higgs field has a self–coupling of order one, then the mass of the Higgs boson is of order v, and it is appropriate to integrate out the Higgs boson so that it does not appear as one of the explicit degrees of freedom in the effective theory. Similarly if some of the fermions have order–one Yukawa couplings to the Higgs field, then they too acquire masses of order v, and should be integrated out. It is possible that the fermions that acquire large masses are in representations of the gauge group that give a nontrivial contribution to the gauge anomaly; then the remaining light fermions in the effective field theory have uncanceled gauge anomalies. The effective field theory is thus a theory of massive gauge fields coupled to fermions in a representation of the gauge group with uncanceled gauge anomalies. Yet this theory is surely unitary and Lorentz invariant, because it merely describes the low–energy physics of an underlying renormalizable gauge theory that is known to be consistent. On the other hand, this effective field theory is surely not renormalizable; it breaks down as a description of the physics of

our model for processes with typical energies of order v or greater. We have, at any rate, a constructive proof of the existence of a consistent (but nonrenormalizable) effective field theory of (massive) gauge fields coupled to fermions in an anomalous representation.

Furthermore, we have no sound reason to be disappointed that the gauge fields in our effective field theory turn out to be massive. The only reason to expect the gauge fields to be massless is that masses are forbidden by gauge invariance. But if our goal is to construct a consistent effective field theory that has uncanceled gauge anomalies, then we ought not to expect massless gauge fields, because gauge invariance is broken by the anomaly. Indeed, even if the bare gauge field masses vanish in an anomalous effective field theory, we should expect nonvanishing gauge field masses to be generated by radiative corrections.[★]

We recognize, then, that anomalous gauge theories (in four spacetime dimensions) can certainly be quantized, if we regard as an acceptable quantum theory an effective field theory with an explicit cutoff Λ, and in which the gauge field has a nonzero mass μ. It is now possible to formulate the central question that I intend to address in this paper: Given an anomalous effective gauge field theory with cutoff Λ, how high can Λ be pushed up before the effective field theory description breaks down and new physics must be encountered? This is the question that we would want to answer in a phenomenological context, if we discovered that an anomalous gauge theory provided a good description of low–energy physics. Just as the nonrenormalizability of Fermi's weak interaction theory indicated that a new description would be needed at high energy, the anomaly would be regarded as an indication of new physics awaiting discovery at higher energies. Naturally, we would want to know how inaccessible that new physics might be.

★ This statement is not precisely correct as it stands, because unrenormalized perturbation theory cannot be sensibly formulated in an anomalous gauge theory unless the bare gauge field masses are nonzero. The correct statement is that, even if bare gauge field masses are chosen to be infinitesimally small, the renormalized masses will assume finite values due to large radiative corrections. See Section 2.

What criterion should we invoke to determine that the effective field theory description of the physics is breaking down? The cutoff Λ in the effective field theory plays a dual role; it is both the maximum energy of the processes described by the effective theory, and the maximum wave number of the quantum fluctuations described by the theory. In terms of Feynman diagrams, Λ in the former role is a limit on the external momenta carried by the diagrams, and in the latter role is a cutoff on the loop momenta in the diagrams. In its role as a cutoff on loop momenta, Λ controls the size of radiative corrections, and in particular determines the size of coupling renormalization. We will take the attitude that the effective field theory loses all predictive power, and must be replaced by a more complete description, if a process at the maximum energy Λ can be calculated in the effective theory only in terms of an infinite number of free renormalized parameters, or if it can be calculated in terms of a finite number of renormalized parameters only if an infinite number of bare parameters are finely "tuned." Thus, *calculability* is the fundamental test that must be met by a sensible effective field theory.

The criterion annunciated in the previous paragraph was formulated in the language of renormalized perturbation theory. The physical meaning of this criterion is clarified if we restate it in a nonperturbative language. We may contemplate calculating the exact renormalization group flow (in the sense of Wilson[3-5]) of an anomalous gauge theory, even though such a calculation may be beyond our analytical ability in practice. That is, we may allow the cutoff Λ on loop momenta to "float" upward, and meanwhile modify the action S_Λ of the theory so that the physics at a fixed energy scale well below Λ is unchanged. Typically, for a theory that is perturbatively nonrenormalizable, S_Λ must eventually, as Λ increases, cease to be "quasilocal" in spacetime on distance scales comparable to Λ^{-1}. There is thus a maximum value $\Lambda_{\max}$ of the floating cutoff; at distances smaller than $\Lambda_{\max}^{-1}$, the locality of the theory necessarily breaks down.

This failure of locality at sufficiently short distances means that amplitudes at sufficiently high energy have a peculiar nonanalytic dependence on external momenta. Physically, this nonanalyticity represents a failure of unitarity, because unitarity re-

lates the singularities of amplitudes to physical intermediate states. To restore unitarity, then, if this is possible at all, we must introduce into the theory new massive quanta with masses of order Λ_{max}. By generating an infinite number of counterterms, perturbation theory does its best to signal the breakdown of locality and hence of unitarity. Perturbation theory in a nonrenormalizable theory is an expansion in powers of external momenta, an expansion that necessarily breaks down in a nonlocal theory. Therefore, the energy scale at which perturbative calculability is lost provides at least an order–of–magnitude estimate of the limiting cutoff Λ_{max}.

The question concerning the maximal allowed value of the cutoff Λ in an anomalous gauge theory with renormalized gauge boson mass μ has an obvious answer that is suggested by our "constructive proof" of the existence of quantized anomalous gauge theories. It is reasonable to expect that $\Lambda \lesssim \mu/e$ is required (where e is the gauge coupling), because that is the mass scale at which we expect the Higgs boson and the massive fermions to appear. We will see that this answer is correct in the case of an anomalous non–Abelian gauge theory, but that the answer in the Abelian case is quite different. This distinction between non–Abelian and Abelian theories arises because massive Yang-Mills theory is nonrenormalizable even in the absence of anomalies, while a massive anomaly-free Abelian gauge theory is renormalizable. From the perspective of an underlying renormalizable theory, the distinction can be traced to the fact that charge is not quantized in the Abelian case. At any rate, it is instructive to see that this question can be answered by analyzing the effective field theory itself, without considering any candidate underlying renormalizable theories that reduce to the effective field theory in question at low energy.

Before proceeding to the analysis, I would like to comment on the connection between my point of view and an alternative approach to the quantization of anomalous four–dimensional gauge theories. When one is confronted with an anomalous gauge theory, one feels a strong temptation to cancel the anomaly and restore gauge invariance. Of course, the conventional way to do this is to introduce additional fermions in representations chosen appropriately to cancel the anomaly; one then obtains an anomaly–free renormalizable theory. But one might also contemplate restoring gauge

invariance in an unconventional way, by introducing other new degrees of freedom with suitable couplings and gauge transformation properties, at the potential cost of losing power–counting renormalizability.

There is indeed an alternative way to cancel the anomaly in a four–dimensional gauge theory, without introducing new fermions. In this procedure[6,7], gauge degrees of freedom are elevated to the status of dynamical scalar fields, and a Wess–Zumino counterterm[8,9] is constructed from scalar and gauge fields that compensates for the gauge anomalies of the fermions. However, merely "canceling the anomaly" with such a Wess–Zumino term actually does nothing at all by itself to cure the disease of the original anomalous theory. The new "dynamical" variables are just gauge artifacts that can be completely removed by appropriate gauge fixing.[★] So when we introduce fictitious scalar fields and a Wess–Zumino term in order to cancel the gauge anomalies of the fermions, we are only erecting a gauge–invariant *description* of the original anomalous theory, and we can recover the original description by choosing the "unitary" gauge. In this gauge–invariant description, gauge field masses may be regarded as arising from the Higgs mechanism; the fictitious scalars are the Goldstone bosons that get "eaten" by the gauge bosons. Thus, the observation that nonvanishing gauge field masses are necessary in the anomalous theory can be restated, in the gauge–invariant language, by saying that the local symmetry of an anomalous theory must be nonlinearly realized.

The gauge–invariant description turns out to be quite convenient, and we will use it in our analysis. When we have a gauge–invariant formulation, we are free to power–count Feynman diagrams in any gauge we choose, and it is simpler to perform the power counting in an "'t Hooft gauge," with a well–behaved gauge–field propagator, than in a "unitary gauge," with a poorly–behaved gauge–field propagator. The gauge–invariant description also allows us to easily discern nontrivial Ward identities that survive in spite of the anomaly; these identities can be derived as consequences of the BRST[10] symmetry of the 't Hooft–gauge version of the theory.

★ At least, this is true *locally*. Considerations of global topology will be mostly ignored in this paper.

The remainder of this paper is organized as follows: in Section 2, we derive an upper bound on the cutoff Λ in an anomalous *Abelian* effective gauge theory, and in Section 3 we explain how this bound can be saturated by an effective theory obtained as the low–energy limit of an underlying renormalizable theory. We extend our results to theories with mixed Abelian–non–Abelian anomalies, including gravitational anomalies, in Section 4. Non–Abelian gauge anomalies are considered in Sections 5 and 6. We describe in Section 5 a procedure for quantizing an anomalous theory with non–Abelian gauge group G in which global G symmetry is maintained; Section 6 concerns an alternative procedure in which an anomaly–free subgroup H of G remains linearly realized. Global anomalies are the subject of Section 7, and local discrete symmetries are considered in Section 8. Section 9 contains our conclusions, including some remarks about the "phenomenological" implications of this paper.

There is now a substantial literature concerning the problem of quantizing anomalous gauge theories. Many authors, beginning with Jackiw and Rajaraman[11] have restricted their attention to the case of two–dimensional spacetime, for which the issue of renormalizability has a much different character than in four dimensions. Two-dimensional theories will not be explicitly discussed in this paper. The proposal that a four-dimensional anomalous gauge theory can be sensibly quantized by introducing an appropriate Wess–Zumino counterterm was pioneered by Faddeev and Shatashvili[6] and Krasnikov[7], and further elaborated by Babelon, Shaposnik, and Viallet[12] and by Harada and Tsutsui[13]. Since then, various explicit constructions of four-dimensional anomalous gauge theories have been described[14-23]. This previous work overlaps considerably with the analysis of the Abelian case described in Section 2 of this paper. However, most previous authors have either chosen not to address the issue of renormalizability, or have reached different conclusions than mine.[†]

Nevertheless, many aspects of the work reported here have been anticipated by

---

† Aoki[23] concludes that an anomalous Abelian gauge theory is nonrenormalizable, but does not attempt to obtain a bound on the cutoff. That anomalies induce a failure of renormalizability was also the conclusion of the classic work in [1].

other authors. In particular, the central thesis of this paper — that any sensible method of quantizing an anomalous gauge theory in four dimensions inevitably results in a nonrenormalizable theory with nonlinearly realized gauge symmetry — has also been forcefully advocated by Einhorn and Goldberg [24] and by Ball [25]. And the idea that an anomalous effective gauge theory can be constructed by integrating out heavy fermions from an underlying anomaly–free theory was suggested by Sterling and Veltman [26], and definitively discussed by D'Hoker and Farhi [27].

2. A Chiral Abelian Gauge Theory[‡]

In this section we will examine the simplest example of a four–dimensional gauge theory that suffers from a chiral gauge anomaly. This is a theory of a single left–handed Weyl fermion ψ_L (with charge Q) coupled to a U(1) gauge field A. The Lagrange density of this theory is

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi}_L \gamma^\mu i(\partial_\mu - ieQA_\mu)\psi_L \quad . \tag{2.1}$$

An effective action $\Gamma[A]$ for the gauge field can be constructed by "integrating out" the fermion ψ_L; it is defined by[§]

$$e^{i\Gamma[A]} = \int d\psi_L d\bar{\psi}_L \, e^{i \int d^4x \mathcal{L}(\psi_L, \, \bar{\psi}_L, \, A)} \quad . \tag{2.2}$$

The statement that this theory has a chiral gauge anomaly means that $\Gamma[A]$ is not a gauge–invariant functional of A. Indeed, under a gauge transformation

$$A_\mu \to A_\mu + \frac{1}{e} \partial_\mu \omega \quad , \tag{2.3}$$

---

‡ The work described in this section was done in collaboration with Mark Wise[28].
§ Actually, it is a delicate matter to give a well-defined meaning to this formal expression; see, for example, [29].

Γ changes according to[¶]

$$\delta_\omega \Gamma \;=\; \int \, d^4 x \; \frac{e^2}{48\pi^2} \; Q^3 \omega \; F_{\mu\nu} \tilde{F}^{\mu\nu} \quad , \tag{2.4}$$

where $\tilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\lambda\delta}F^{\lambda\delta}$ is the dual of F.

We propose to study the perturbation expansion of this quantum theory, regarding it as an effective field theory with an explicit ultraviolet cutoff Λ. For this purpose, we will find it convenient to use a technical trick that is a variant of a similar trick introduced long ago by Stueckelberg in a slightly different context. We will introduce into our mathematical description of this theory, without modifying the physics in any way, a new degree of freedom that is "pure gauge." This new degree of freedom is purely fictitious, because the modified theory has a gauge invariance that allows us to remove the new degree of freedom by a suitable choice of gauge. Nonetheless, we will find this modified description very convenient; by a felicitous choice of gauge, we will be able to greatly simplify the power counting of the Feynman diagrams that arise in the perturbation expansion.

The new degree of freedom that we will introduce is a dimensionless scalar field $\theta(x)$ that transforms under a gauge transformation as

$$\theta \to \theta \,+\, \omega \quad . \tag{2.5}$$

We wish to couple θ to the gauge field A in such a way that gauge invariance at the quantum level is restored. Evidently, then, we should add to our Lagrange density the term

$$\mathcal{L}_\theta \;=\; -\,\frac{e^2 Q^3}{48\pi^2} \; \theta \; F_{\mu\nu} \, \tilde{F}^{\mu\nu} \quad ; \tag{2.6}$$

the variation of this contribution to the action under a gauge transformation cancels the anomalous contribution eq. (2.4). If one wishes, $\mathcal{L}_\theta$ can be put in an alternative

---

[¶] This transformation property can be inferred from the calculation in [30] of the diagrams shown in Fig. 1. There is, however, a factor of $\frac{1}{3}$ that distinguishes this "consistent" form of the anomaly from the "covariant" form[31,29].

form by performing the change of variable

$$\psi_L \to e^{-iQ\theta} \; \psi_L \; , \; \bar{\psi}_L \to e^{iQ\theta} \; \bar{\psi}_L \quad , \tag{2.7}$$

which has a nontrivial Jacobian[32] that cancels eq. (2.6).* We then obtain

$$\mathcal{L}_\theta \; = \; Q\partial_\mu\theta\bar{\psi}_L\gamma^\mu\psi_L \quad . \tag{2.8}$$

Eq. (2.6) can be recovered from eq. (2.8) by an integration by parts, if we recall that the current $J^\mu = \bar{\psi}_L\gamma^\mu\psi_L$ is not conserved, due to the anomaly. The expression for $\mathcal{L}_\theta$ in eq. (2.8) serves to remind us that, despite the appearance of eq. (2.6), the field θ is in fact derivatively coupled.

Now that we have constructed the action of our theory, we must explain how it is to be quantized; we will settle for a prescription that makes sense to any finite order of perturbation theory. Formally, we will define the quantum theory by the Faddeev–Popov[33] path integral method. This means that we will impose a gauge condition and integrate over the dynamical variables ψ, $\bar{\psi}$, A, θ on a "gauge slice," including the functional determinant that ensures that the results are independent of the choice of the slice.

It is clear, then, that the quantum theory that we obtain from our modified description in which θ is a dynamical variable is equivalent, at least formally, to the quantum theory that we would have obtained had we integrated over ψ, $\bar{\psi}$, A without introducing θ. By a suitable gauge transformation, we may impose the gauge condition

$$\theta(x) \; = \; 0 \quad . \tag{2.9}$$

The integration is restricted to this gauge slice if we introduce the δ–functional $\delta[\theta(x)]$ into the functional integral. The associated Faddeev–Popov determinant is trivial.

$\star$ The Jacobian must be regulated with care in order to get the normalization right. See [29] for details.

We are left then, with an integral over the remaining variables ψ, $\bar{\psi}$, A, weighted by $\exp(i \int d^4x \mathcal{L})$ where $\mathcal{L}$ is the Lagrange density defined in eq. (2.1). (Because the gauge field is necessarily massive, as we will see below, eq. (2.9) suffices to fix the gauge completely.)

The above discussion is merely formal because we have made no mention of the regulator that is needed to define the path integral. We will not address here the problem of specifying a nonperturbative regulator; we instead regard the path integral as an object that generates the Feynman diagram expansion. For our discussion we will not really need to specify exactly how the diagrams are regulated, but it is natural to consider all integrations over loop momenta in diagrams to be cut off at a momentum scale Λ. Such a sloppy method of regularization brutally breaks the gauge invariance of our theory, but this is not a great loss. Because of the anomaly, no gauge–invariant regulator exists anyway. Of course, to ensure that our quantum theory is independent of the gauge slice, we must add local counterterms to the action that restore gauge invariance order–by–order in perturbation theory.

With a regulator in place, and with the understanding that local counterterms are to be added order–by–order to restore gauge invariance, our formal argument presented above shows that the quantum theory that we have constructed is unitary, at least in perturbation theory, if all external momenta are small compared to Λ. Unitarity holds because our regulated Feynman rules satisfy cutting rules, and the gauge $\theta = 0$ is a "unitary" gauge in which all propagating degrees of freedom correspond to physical particles. A reader who is skeptical about the alleged unitarity of this theory will be reassured when reminded that it can be obtained as the low–energy limit of an underlying renormalizable gauge theory, as we will discuss explicitly in Section 3.

Incidentally, if we try to extend beyond perturbation theory our demonstration of the equivalence of the theories with and without the field $\theta(x)$, we confront a subtlety associated with the topology of the θ field. As we may regard the term eq. (2.6) as the Jacobian of the transformation eq. (2.7), it is natural to interpret θ as an angular variable with period $2\pi/Q$. But if θ is an angular variable, then, although we can

always impose the gauge condition $\theta(x) = 0$ locally in spacetime, there may be a topological obstruction to imposing it globally, for there may be vortex excitations around which θ has a nontrivial winding number. Thus, it seems that the two theories can be distinguished by the criterion of the existence of topologically stable vortex excitations. We will ignore the vortices, however, because they have no effect on the perturbative behavior that we wish to study. In any event, the vortices have a size of order Λ^{-1}, and since they are associated with physics at the scale of the short distance cutoff, it is presumably sensible to ignore them when we apply our effective field theory to physics at energies well below Λ. The formal equivalence of the two formulations of the theory can be restored if we regard θ as noncompact — that is, we integrate only over the field configurations such that $\theta(x)$ is single–valued.[†]

Before we proceed with the perturbative analysis of this theory, it may be worth-while to emphasize again the interpretation of the modification to the anomalous gauge theory that we have made by introducing the θ field. Since the new term, eq. (2.6), compensates for the gauge anomaly of the Weyl fermion, one is tempted to say that we have "canceled" the anomaly by introducing a new physical degree of freedom, just as we might cancel the anomaly by introducing a right–handed Weyl fermion with charge Q. Indeed, eq. (2.6) is the four–dimensional analog of the counterterm that Green and Schwarz[34] introduced into ten–dimensional physics to effect anomaly cancellation (and it arises in some four–dimensional compactifications of superstring theory[35]).[‡] But it is quite misleading to interpret our construction as a mechanism for anomaly *cancellation*. Since the field θ that we have introduced is a gauge artifact, we have merely found a new and more convenient *description* of the physics of the anomalous gauge theory.

The reason that our new description is more convenient is that the original description can now be regarded as a particular choice of gauge in the modified theory.

[†] There exists no comparable procedure for removing topological excitations, however, in the non–Abelian gauge theories that we will consider later. For the most part, we will ignore topological subtleties in this paper.

[‡] An effective field theory description of this "anomaly cancellation" mechanism was presented in[36].

And this choice of gauge — the unitary gauge $\theta(x) = 0$ — is not at all the most efficient for doing explicit computations. In particular, power counting in the unitary gauge is subtle, because of delicate cancellations that occur among diagrams. It is much better to choose an 't Hooft gauge[2], in which power counting is simple.[§] For us, the most natural choice will be the 't Hooft–Landau gauge, in which the field θ remains derivatively coupled.

To see how to implement the 't Hooft gauge condition, one should first appreciate that our effective field theory really describes a U(1) gauge theory in a Higgs phase. This is more or less obvious, since the gauge symmetry $\theta \rightarrow \theta + \omega$ is evidently spontaneously broken in the ground state with $\theta = $ constant. Of course, to verify that the gauge symmetry is realized via the Higgs phenomenon, we should go to unitary gauge and check that the photon is massive.

First, we note that, although we have included no kinetic term for θ in our bare Lagrangian, such a term will be generated as a radiative correction. When the external momentum p of the θ field is small, the diagram shown in Fig. 2 is, in order of magnitude,[¶]

$$\text{Diagram} \ \sim \ \left(\frac{1}{16\pi^2} \right) \ \left(\frac{e^2 Q^3}{16\pi^2} \right)^2 \ p^2 \Lambda^2 \quad . \tag{2.10}$$

There is no constant term independent of p because of gauge invariance; hence the diagram is quadratically divergent by power counting. When combined with the diagrams of Fig. 3, the diagram of Fig. 2 generates for θ the gauge–invariant kinetic term

$$\mathcal{L}_{\text{kin}} \ = \ \frac{1}{2} \ F^2 \ (\partial_\mu \theta \ - \ e A_\mu)^2 \tag{2.11}$$

where

$$F \ \sim \ \frac{e^2 |Q^3|}{64\pi^3} \ \Lambda \quad . \tag{2.12}$$

In the unitary gauge $\theta(x) = 0$, this kinetic term reduces to a mass term for the gauge

---

§ Fictitious local symmetries are introduced with a similar motivation in, for example, [37].

¶ I am being rather cavalier with numerical factors; 4/3 has been replaced by 1 here.

field A, with

$$\mu = eF \quad, \tag{2.13}$$

and F may be interpreted as a symmetry–breaking mass scale. It is an old observation[1], of course, that the anomaly through the diagram in Fig. 3b generates a photon mass. Naturally, if we introduce the gauge artifact θ, we can interpret the photon mass as arising because the θ particle is "eaten" by the photon.

Actually, the discussion of the preceding paragraph requires some clarification. In the anomalous gauge theory formulated as in eq. (2.1), it is inappropriate to "fix the gauge," because gauge invariance is broken by the anomaly. (Correspondingly, in the gauge–invariant formulation of the theory, eq. (2.9) should suffice to fix the gauge completely.) But if we sum over all gauge field configurations in the path integral, without any gauge fixing, then the perturbative propagator for a massless gauge field is ill–defined. In order that the path integral make sense, we should consider the bare gauge field mass to be infinitesimal rather than zero. Then the loop corrections *amplify* the mass to the finite value given by eq. (2.12, 2.13).

If the kinetic term for θ is of the form eq. (2.11), it is convenient to introduce a rescaled field

$$b = F\theta \quad, \tag{2.14}$$

that has a conventionally normalized kinetic term,

$$\mathcal{L}_{\text{kin}} = \frac{1}{2} \left(\partial_\mu b - \mu A_\mu \right)^2 \quad. \tag{2.15}$$

Then the choice

$$\mathcal{L}_{\text{g.f.}} = -\frac{1}{2\alpha} \left(\partial_\mu A^\mu - \alpha \mu b \right)^2 \quad, \tag{2.16}$$

for the gauge fixing term removes the $A - b$ mixing. These are the covariant 't Hooft gauges, parametrized by α. In these gauges, the photon propagator has good ultraviolet behavior, and the b has mass $m^2 = \alpha \mu^2$; there is no need for a Faddeev–Popov

ghost. Particularly convenient is the case $\alpha = 0$, the 't Hooft–Landau gauge, in which b is massless. The unitary gauge is recovered in the limit $\alpha \to \infty$, as in this limit the b particle is removed from the spectrum (and the nice ultraviolet behavior of the photon propagator is lost). The usual arguments[2] based on gauge invariance can be invoked to show that all physics is independent of α.

The estimate of F in eq. (2.12) is the *natural* value for F in the renormalized Lagrangian of our effective theory; a much smaller value of F could be realized only by carefully tuning the bare value of F so that it nearly cancels the radiative correction. Since F may be interpreted as a symmetry breaking mass scale, it is somewhat of a surprise to find that this natural value for F can be much smaller than the cutoff Λ if the coupling is weak. One is inclined to interpret our effective field theory as the low–energy limit of an underlying renormalizable gauge theory that has undergone the Higgs mechanism, and hence to expect the physics associated with the symmetry breakdown (the physical Higgs particles and massive fermions that cancel the anomaly) to have a characteristic mass scale above or of the order of Λ. It seems odd, then, that the apparent symmetry breaking mass scale F can be much smaller than Λ. In Section 3 we will see explicitly what type of physics at the scale Λ can produce a photon mass so much smaller than Λ. (The key observation is that electric charge is unquantized in an Abelian gauge theory, and thus the Higgs field can in principle carry an arbitrarily small charge.)

But for now, let us explore the question whether the *natural* value for F in eq. (2.12) is really a *minimal* value. That is, we will investigate how small μ/Λ can be in our effective theory, where μ is the photon mass and Λ is the cutoff, if we demand that amplitudes for processes with characteristic energies of order Λ can be computed in terms of a finite number of renormalized parameters. To appreciate what is at issue, imagine that it were possible, by tuning one or a finite number of parameters, to arrange for the ratio μ/Λ to be arbitrarily small. Then we would say that an anomalous Abelian gauge theory, while failing the criterion of naturalness, is renormalizable. It would have a status similar to that of a theory of a self–coupled elementary scalar field[38,39].

In fact, our anomalous Abelian gauge theory is not renormalizable. The nonrenormalizability arises in the unitary gauge because of the bad ultraviolet behavior of the photon propagator. But it can be understood more directly in the 't Hooft–Landau gauge as arising because the term, eq. (2.6) in the Lagrange density, is a dimension–5 operator (when expressed in terms of the conventionally normalized scalar field b). The terms in the Lagrange density involving the field b are

$$\mathcal{L} = \frac{1}{2} \left(\partial_\mu b - \mu A_\mu \right)^2 - \frac{1}{3v} \, b F_{\mu\nu} \tilde{F}^{\mu\nu} \qquad (2.17)$$

where b transforms under a gauge transformation as

$$b \to b + (\mu/e) \, \omega \qquad (2.18)$$

and gauge invariance requires

$$\frac{1}{v} = \frac{(eQ)^3}{16\pi^2 \mu} \quad . \qquad (2.19)$$

The naturalness criterion of eq. (2.12) may be restated as

$$\Lambda \lesssim 4\pi |v| \quad . \qquad (2.20)$$

But standard power–counting arguments, which can be justified in the 't Hooft–Landau gauge, show that because the coupling constant v^{-1} has the dimensions of inverse mass, radiative corrections generate counterterms of arbitrarily high dimension. And furthermore, if Λ is chosen to violate eq. (2.20), then these counterterms give non–negligible contributions to the amplitudes for processes with typical energies of order Λ. (The factor 4π in eq. (2.20) arises because of the $1/16\pi^2$ that accompanies each loop in the Feynman diagram expansion.) Thus, eq. (2.20) is more than just a naturalness requirement; it is the condition for *calculability* in our effective field theory. When eq. (2.20) is not satisfied, nothing can be computed in terms of a finite

16

number of parameters, unless we are willing to tune an *infinite* number of parameters to unnaturally small values. We conclude, then, that the condition

$$\Lambda \lesssim \mu / \left(|eQ|^3 / 64\pi^3 \right) \tag{2.21}$$

is a fundamental limitation on the cutoff of an anomalous Abelian gauge theory (in four dimensions).

Since our power–counting argument is conventional, we need not go through it in detail; it will suffice to briefly sketch the argument. First recall that the diagrams shown in Figs. 2 and 3 generate a local operator of order

$$\left(\frac{\Lambda}{4\pi v} \right)^2 (\partial b - \mu A)^2 \quad . \tag{2.22}$$

Thus, the b field and the photon mass μ are renormalized by a large factor if $\Lambda >> 4\pi v$. This is the "naturalness" problem referred to earlier; the large renormalization of μ is avoided only if the bare mass of the photon is carefully tuned.

Similarly, the diagrams shown in Fig. 4 generate the local operator

$$\frac{1}{(4\pi v)^2 v^2} (\partial b - \mu A)^4 \tag{2.23}$$

ignoring a logarithm of the cutoff). If the four–point coupling for b in eq. (2.23) is inserted in the diagram of Fig. 5, then this diagram, and others that form with it a gauge–invariant set, generate the operator

$$\left(\frac{\Lambda}{4\pi v} \right)^4 (\partial b - \mu A)^2 \quad . \tag{2.24}$$

As we continue to iterate the loop expansion, higher and higher powers of $\Lambda/4\pi v$ occur in the renormalization of b and μ.

In fact, all of the other local operators generated by the loop expansion behave the same way. For example, the diagram of Fig. 6 generates the local operator

$$\left(\frac{\Lambda}{4\pi v}\right)^4 \frac{1}{(4\pi v)^4 v^4} (\partial b - \mu A)^6 \quad ; \tag{2.25}$$

and if the six–point coupling for b in eq. (2.25) is inserted in the diagram of Fig. 7, we obtain

$$\left(\frac{\Lambda}{4\pi v}\right)^8 \frac{1}{(4\pi v)^2 v^2} (\partial b - \mu A)^4 \quad . \tag{2.26}$$

As we continue to iterate the loop expansion, arbitrarily high powers of $\Lambda/4\pi v$ are generated, multiplying all of the local operators

$$\frac{1}{(4\pi v)^{p-2}} \frac{1}{v^{p-2}} (\partial b - \mu A)^p \quad . \tag{2.27}$$

Therefore, for $\Lambda \gg 4\pi v$, the perturbative loop expansion breaks down grievously, and the amplitudes for processes with characteristic energies of order Λ receive non–negligible contributions from operators of all dimensions (unless the coefficients in the bare action of all but a finite number of operators are tuned to arbitrary accuracy). We have no choice but to conclude that our effective field theory description of the physics has broken down, and must be replaced by a more complete description at energy scales exceeding $4\pi v$. Consistency, then, requires that the cutoff Λ in the effective field theory satisfy the bound eq. (2.21).

As we noted in the introduction, we may also think of the maximal value of Λ as the inverse distance scale at which the effective field theory ceases to be quasilocal. Of course, the crude perturbative analysis described here gives at best an order of magnitude estimate of the actual upper bound on Λ. To do much better would require much more sophisticated nonperturbative methods.

To avoid misunderstanding, I should perhaps remark that our bound, eq. (2.21), would also apply to a supersymmetric generalization of our anomalous gauge theory. The Λ–dependence revealed by our power–counting analysis will not be tamed by

supersymmetry. The local operators discussed above, suitably extended to a super-symmetric form, are not protected by nonrenormalization theorems.

3. Saturating the Bound

The analysis of the previous section showed that an anomalous Abelian gauge theory with an explicit cutoff Λ is consistent if the photon mass μ satisfies the upper bound

$$\mu \gtrsim \frac{|eQ|^3}{64\pi^3} \Lambda \quad . \tag{3.1}$$

(Of course, if the theory contains many fermion species, Q^3 should be replaced by $\mathrm{tr}(Q^3) = \sum_i Q_i^3$, summed over fermion species.)

The existence of this consistent effective field theory suggests that it ought to be possible to construct an underlying renormalizable (anomaly–free) gauge theory that reduces to our effective field theory below the mass scale Λ. In this section we will describe how such an underlying renormalizable theory can be constructed. This construction shows that the bound eq. (3.1) really is optimal, or in other words, that it can be saturated.

The construction takes advantage of the fact that electric charge is not quantized in a U(1) gauge theory. Our underlying theory may be chosen to be an Abelian Higgs model in which the Higgs field ϕ has a charge q that is as small as we please. If ϕ, a conventionally normalized complex scalar field has vacuum expectation value η, then we may express

$$\phi = \frac{\eta}{\sqrt{2}} e^{iq\theta} + \dots \quad , \tag{3.2}$$

and the gauge–invariant kinetic term for ϕ becomes

$$\mathcal{L}_{\mathrm{kin}} = |(\partial_\mu - ieA_\mu)\phi|^2 = \frac{1}{2} q^2\eta^2(\partial\theta - eA)^2 + \dots \quad , \tag{3.3}$$

where θ is the Goldstone boson field that is "eaten" by the photon. Comparing to

19

eq. (2.11), we see that $F = q\eta$ is the "symmetry–breaking scale" and

$$\mu = eq\eta \qquad (3.4)$$

is the photon mass, in the tree approximation to the underlying theory.

Let us suppose, at first, that our renormalizable U(1) gauge theory contains no fermions, or that it contains fermions in anomaly–free representations that will be retained in the low–energy effective field theory. Then we may choose the mass of the physical Higgs particle (the modulus of ϕ) to be of order η; this corresponds to choosing the Higgs self–coupling to be of order one. By integrating out the physical Higgs boson, we obtain an effective field theory with cutoff $\Lambda \sim \eta$ that contains a photon of mass $\mu \sim eq\Lambda$. By choosing q small we may make the photon mass arbitrarily small compared to the cutoff Λ. This is an easy way to see that a theory of a massive U(1) gauge boson is renormalizable, if the gauge boson couples to a conserved current.

Now imagine that our renormalizable U(1) gauge theory contains several species of charged Weyl fermions, with canceling anomalies. It is possible that some of the fermions have Yukawa couplings to the Higgs scalar ϕ while others do not. (This is natural in the technical sense if the U(1) charges of the uncoupled fermions are such that a gauge–invariant dimension–four Yukawa coupling cannot be constructed.) If the Yukawa couplings are of order one, these fermions acquire masses of order η, and disappear from the effective field theory that describes physics below the cutoff $\Lambda \sim \eta$. If the fermions that acquire large masses have charges such that

$$\sum_{\text{heavy}} Q_i^3 \neq 0 \quad , \qquad (3.5)$$

then the charges of the remaining light fermions satisfy

$$\sum_{\text{light}} Q_i^3 = -\sum_{\text{heavy}} Q_i^3 \neq 0 \qquad (3.6)$$

(because the underlying theory is anomaly–free), and the effective field theory is an

20

anomalous Abelian gauge theory. (If $eq << 1$, then the photon, though massive, is much lighter than the cutoff, and is retained in the effective field theory.)

There is one minor obstacle to this strategy. If the Higgs field ϕ has a charge q, then Weyl fermions ψ_1, ψ_2 can be Yukawa coupled to ϕ,

$$\mathcal{L}_{\text{Yuk}} = -\lambda \, \psi_1 \psi_2 \phi + \text{h.c.} \quad , \tag{3.7}$$

only if the charges $Q_{1,2}$ of the fermions satisfy

$$Q_1 + Q_2 = -q \quad . \tag{3.8}$$

Then the fermions ψ_1, ψ_2 give a contribution to the anomaly that is proportional to

$$Q_1^3 + Q_2^3 = Q_1^3 - (Q_1 + q)^3 \sim -3Q_1^2 q \quad , \tag{3.9}$$

which is of order q. Therefore, if $q << 1$, then the pair of fermions ψ_1, ψ_2 that acquire large mass when ϕ acquires a vacuum expectation value give a small contribution to the anomaly, and the remaining light fermions in the effective field theory must also have a small anomaly.

It is easy to overcome this objection. For example, nothing prevents us from including in the underlying renormalizable theory a number of order $\frac{1}{q}$ of fermion species, all of which are Yukawa coupled to ϕ, and all of which are integrated out when we construct the effective field theory. Each species gives a small contribution to the anomaly, but their accumulated effect can be order one. Thus, the anomaly of the light fermions that remain in the effective theory need not be small.

Now we have seen how to construct an underlying renormalizable U(1) gauge theory that reduces below a cutoff $\Lambda \sim \eta$ to an anomalous U(1) gauge theory. Furthermore, we may choose the charge q of the Higgs field so that the *tree approximation* contribution to the photon mass, $\mu = eq\eta$, is as small as we please. But we must also consider the effect of loop diagrams that contribute to the photon mass. For example,

the photon mass is renormalized by the one–loop diagram in Fig. 8, if the fermion appearing in the loop has acquired a non-gauge-invariant mass through its Yukawa coupling to the Higgs field. However, when the diagrams in Fig. 8 and Fig. 9 are combined, we recognize that these diagrams merely rescale the gauge–invariant kinetic term eq. (3.3) of the Higgs field; they do not modify the tree–approximation relation $\mu = eq\eta$, if μ is interpreted as the renormalized photon mass, e is the renormalized charge, and η is the expectation value of the renormalized Higgs field.

The contribution to μ that does significantly modify this tree–approximation relation comes from the diagram in Fig. 3b. From the point of view of the underlying theory, the fermions running around the triangles in this diagram can be either heavy fermions or massless fermions. If all fermions were massless, the contributions to μ from the diagram in Fig. 3b would cancel, when all fermion species were summed over. But if some of the fermions have large mass M, then this cancellation is incomplete. An estimate of the contribution to μ due to this diagram is most conveniently done by invoking the effective field theory viewpoint. Beginning with the underlying renormalizable theory, which has an infinite cutoff, we imagine gradually lowering the cutoff by integrating out short–wavelength degrees of freedom. When we have reduced the cutoff to of order M, the photon mass that has been induced is negligible, because of the cancellation between the effects of heavy and light fermion species. Then we integrate out the heavy fermions, and the remaining contribution to μ is that computed in Section 2,

$$\mu^2 \sim \left(\frac{e^3(\text{ tr } Q^3)}{64\pi^3} M \right)^2 \quad , \tag{3.10}$$

with M playing the role of the cutoff Λ. The important feature of eq. (3.10) is that it contains no reference to the small charge q of the Higgs field; rather, the size of this contribution to μ is controlled by the strength of the anomaly of the light fermions, and the mass of the heavy fermions. We see that the photon mass in this underlying renormalizable theory is consistent with the bound that we derived in Section 2. Indeed, if we choose q to be small enough, the bound is saturated.

Incidentally, it is interesting to note how the coupling eq. (2.6) of the Goldstone boson field θ is generated when we integrate out the heavy fermions to construct the effective field theory. If we write the Higgs field ϕ as in eq. (3.2) then the Yukawa coupling, eq. (3.11) becomes

$$\mathcal{L}_{\text{Yuk}} = -M\, e^{iq\theta}\, \psi_1 \psi_2 + \text{h.c.} \quad , \tag{3.11}$$

where $M = \lambda\eta/\sqrt{2}$. Before we integrate out ψ_1 and ψ_2, it is convenient to perform a change of variable

$$\psi_1 = e^{iQ_1\theta}\psi_1' \quad , \quad \psi_2 = e^{iQ_2\theta}\psi_2' \quad . \tag{3.12}$$

The new fermion fields ψ_1' and ψ_2' are gauge–invariant, and in terms of these variables the fermionic Lagrange density becomes

$$
\begin{aligned}
\mathcal{L}_{\text{fermion}} =\ & \bar{\psi}_1'\, i\gamma^\mu \Big[\partial_\mu + iQ_1(\partial_\mu\theta - eA_\mu)\Big]\psi_1' \\
& + \bar{\psi}_2'\, i\gamma^\mu \Big[\partial_\mu + iQ_2(\partial_\mu\theta - eA_\mu)\Big]\psi_2' \\
& - M(\psi_1'\psi_2' + \psi_1'^\dagger\psi_2'^\dagger) + \text{counterterms} \quad .
\end{aligned}
\tag{3.13}
$$

If we now integrate out ψ_1' and ψ_2', we obviously generate a gauge–invariant functional of $(\partial_\mu\theta - eA_\mu)$. This functional can be expanded in powers of M^{-1}; each term in the expansion is a gauge–invariant local operator.

But there is one local operator generated by integrating out the fermions that is not gauge invariant. The change of variable, eq. (3.12), has a nontrivial Jacobian[32]; it is this Jacobian that gives rise to the operator in eq. (2.6), with Q^3 replaced by

$$Q_1^3 + Q_2^3 \equiv \text{tr}\, Q^3 \quad . \tag{3.14}$$

It is evident that a generalized version of this discussion applies when any number of fermion species are integrated out.

This procedure for integrating out the heavy fermions was studied in detail by D'Hoker and Farhi[27]. Earlier, attention had been called to the role of anomalies when heavy fermions are integrated out by the work of Sterling and Veltman[26].

4. The Mixed Anomaly

If our effective gauge theory has a gauge group that is a product of commuting factors and contains a U(1) factor, then *mixed* anomalies may occur. For the sake of definiteness, suppose that the gauge group is

$$G \; = \; \mathrm{SU}(2) \times \mathrm{U}(1) \quad . \tag{4.1}$$

Then the mixed anomaly is proportional to $\mathcal{A}_R$, where

$$\mathrm{tr}_R(T^a T^b Q) \; = \; \mathcal{A}_R \, \delta^{ab} \quad ; \tag{4.2}$$

here T^a is an SU(2) generator, Q is the U(1) generator, and the trace is taken over the representation R according to which the fermions transform.

If the fermion representation has a nontrivial mixed anomaly, then one might expect that the effective action Γ is not invariant under either SU(2) or U(1) gauge transformations. And one would therefore expect that, in order to construct an operator, analogous to that in eq. (2.6), that can compensate for the change in Γ under both SU(2) and U(1) gauge transformations, we must introduce a scalar field that transforms nontrivially under both SU(2) and U(1).

In fact, however, it is not necessary for the scalar field to transform nontrivially under SU(2). It is possible, by adding to Γ a local counterterm that depends on the gauge fields only, to cancel the variation of Γ under SU(2) gauge transformations. Then the invariance of Γ under U(1) gauge transformations may be restored by coupling the SU(2) gauge fields to a scalar field θ that transforms nontrivially under U(1) only. (This is a four–dimensional variant of the strategy that Green and Schwarz[34] used to deal with mixed anomalies in a ten–dimensional gauge theory.)

To justify the above claim, we investigate how the effective action $\Gamma[A_1, A_2]$ transforms under a gauge transformation. The gauge fields transform as

$$A_1^\mu \rightarrow A_1^\mu + \frac{1}{e_1} \partial^\mu \omega_1 \quad , \tag{4.3}$$

$$A_2^\mu \rightarrow A_2^\mu + \frac{1}{e_2} \partial^\mu \omega_2 + i[\omega_2, A_2^\mu] \quad , \tag{4.4}$$

where e_1 and e_2 are, respectively, the U(1) and SU(2) gauge couplings, and ω_1 and ω_2 are, respectively, infinitesimal U(1) and SU(2) generators. The contribution of the mixed anomaly to the variation of Γ under a gauge transformation may be found by solving the Wess–Zumino consistency condition[8] using, for example, the method of [40]. The result is

$$\delta_{\omega_1}\Gamma = C_1 \frac{e_2^2}{16\pi^2}\mathcal{A}_R \int d^4x \, \omega_1 \left(F_2^{\mu\nu a} \tilde{F}_{2\mu\nu}^a \right) \quad , \tag{4.5}$$

$$\delta_{\omega_2}\Gamma = C_2 \frac{e_1 e_2}{8\pi^2}\mathcal{A}_R \int d^4x \, \tilde{F}_1^{\mu\nu} (\omega_2^a \partial_\mu A_{2\nu}^a) \quad . \tag{4.6}$$

The coefficients C_1 and C_2 are constrained by the strength of the mixed anomaly to satisfy

$$C_1 + C_2 = 1 \quad . \tag{4.7}$$

But the difference $C_1 - C_2$ is *not* constrained. This arbitrariness arises because of the existence of the local counterterm

$$\Gamma_{\mathrm{ct}} = \frac{e_1 e_2^2}{8\pi^2}\mathcal{A}_R \int d^4x \epsilon_{\mu\nu\lambda\sigma} A_1^\mu \left(A_2^{\nu a}\partial^\lambda A_2^{\sigma a} + \frac{1}{3} \, e_2 \, \epsilon_{abc} A_2^{\nu a} A_2^{\lambda b} A_2^{\sigma c} \right) \quad , \tag{4.8}$$

which transforms under a gauge transformation as

$$\delta_{\omega_1}\Gamma_{\mathrm{ct}} = - \frac{e_2^2}{16\pi^2}\mathcal{A}_R \int d^4x \, \omega_1 \left(F_2^{\mu\nu a} \tilde{F}_{2\mu\nu}^a \right) \quad , \tag{4.9}$$

$$\delta_{\omega_2}\Gamma_{\mathrm{ct}} = \frac{e_1 e_2}{8\pi^2}\mathcal{A}_R \int d^4x \tilde{F}_1^{\mu\nu} (\omega_2^a \partial_\mu A_{2\nu}^a) \quad . \tag{4.10}$$

As the Feynman diagrams determine the effective action Γ only up to an arbitrary choice of local counterterms, we are free to redefine Γ by adding to it Γ_{ct} with an

25

arbitrary coefficient. We see that the difference $C_1 - C_2$ in eqs. (4.5) and (4.6) is dependent on the choice of the local counterterm, while the sum $C_1 + C_2$ is independent of this choice.

Evidently, there is a choice of the local counterterm such that $C_2 = 0$, $C_1 = 1$, and Γ is then invariant under SU(2) gauge transformations. The anomalous variation of Γ under a U(1) gauge transformation may then be canceled by a counterterm

$$\mathcal{L}_\theta = -\frac{e_2^2}{16\pi^2}\mathcal{A}_R\theta\left(F_2^{\mu\nu a}\tilde{F}_{2\,\mu\nu}^a\right) \quad , \tag{4.11}$$

where

$$\delta_{\omega_1}\theta = \omega_1 \quad . \tag{4.12}$$

Note that this term has much the same form as the counterterm of eq. (2.6); the difference is that the U(1) gauge field and gauge coupling in eq. (2.6) have been replaced by the SU(2) gauge field and gauge coupling in eq. (4.11), and that the U(1) anomaly $\mathrm{tr}(Q^3)$ has been replaced by the mixed anomaly $\mathrm{tr}(T^aT^bQ)$. Of course, if both $\mathrm{tr}(Q^3)$ and $\mathrm{tr}(T^aT^bQ)$ are nonvanishing, then both of the counterterms of eq. (2.6) and eq. (4.11) would be needed to restore the U(1) gauge invariance of the effective action.

Now, by repeating the power–counting analysis of Section 2, we may derive an inequality constraint on the mass μ of the U(1) gauge boson, namely

$$\mu \gtrsim \frac{e_1e_2^2}{64\pi^3}|\mathcal{A}_R|\,\Lambda \quad , \tag{4.13}$$

where Λ is the cutoff of the effective gauge theory. This bound applies if the SU(2) gauge bosons remain exactly massless; it may be relaxed if the SU(2) gauge bosons are also permitted to acquire mass.

As in the discussion of Section 3, it is possible to construct an underlying (anomaly–free) renormalizable gauge field theory in which the bound eq. (4.13) is saturated. For example, the underlying field theory might contain an SU(2)–singlet

Higgs field ϕ with U(1) charge q, and SU(2)–doublet Weyl fermions ψ_1, ψ_2 with Yukawa couplings to ϕ of the form

$$\mathcal{L}_{\text{Yuk}} \ = \ -\lambda \, \psi_1 \psi_2 \, \phi \ + \ \text{h.c.} \quad . \tag{4.14}$$

This Yukawa coupling is invariant under $G = \text{SU}(2) \times \text{U}(1)$ if the charges Q_1 and Q_2 of the fermion doublets ψ_1 and ψ_2 satisfy $Q_1 + Q_2 + q = 0$. If ϕ acquires a vacuum expectation value, then the $U(1)$ gauge symmetry is spontaneously broken, but the SU(2) gauge symmetry is not. The fermions ψ_1, ψ_2, which carry a mixed anomaly proportional to $Q_1 + Q_2 = -q$, may acquire large masses. If we integrate out the physical Higgs field and the heavy fermions, we obtain an effective field theory with a mixed anomaly; in this effective theory, the SU(2) gauge fields are massless, and the U(1) gauge field is massive. As in Section 3, we may choose q to be small, and choose the number of fermion species that are integrated out to be large. Then the strength $\mathcal{A}_R$ of the mixed anomaly in the effective field theory may be of order one, and the mass of the U(1) gauge boson may saturate the bound eq. (4.13).

Notice that this construction demonstrates that there *must* exist an effective field theory that has both a mixed anomaly and an unbroken SU(2) gauge symmetry. We thus understand better why the counterterm Γ_{ct} with the transformation property eq. (4.10) *had* to exist.

Integrating out the heavy fermions to obtain an effective field theory can be performed as described in Section 3. As in our previous discussion, the change of variable that rotates the phase $e^{iq\theta}$ out of the fermion mass term has a nontrivial Jacobian. It is this Jacobian that generates the term eq. (4.11) in the effective Lagrangian. (For heavy fermions that contribute to the mixed anomaly, the Jacobian involves the SU(2) gauge fields.)

Incidentally, a casual inspection of the diagram in Fig. 10 might cause one to worry that this diagram can generate a mass for the SU(2) gauge boson, if our effective field theory has a nontrivial mixed anomaly. In fact, however, if the appropriate local counterterm is included so as to preserve the SU(2) gauge symmetry, then the

27

diagrams of Fig. 11 cancel this contribution to the gauge boson mass, and the SU(2) gauge bosons remain exactly massless.

In this section, we have described a treatment of the mixed anomaly of an SU(2) × U(1) effective gauge theory in which the SU(2) variation of the effective action is canceled with a local counterterm involving gauge fields only, and the U(1) variation is canceled with a counterterm involving an SU(2)–singlet scalar field that carries a U(1) charge. There is an obvious alternative procedure. We could cancel the U(1) variation of the effective action with a local counterterm involving gauge fields only, and cancel the SU(2) variation with a counterterm involving a scalar field that transforms under SU(2) but not U(1). We will delay a discussion of this alternative until we have investigated the non–Abelian gauge anomaly in Section 5.

It should be evident that the discussion of this section can be generalized to apply to the mixed anomalies of an effective gauge theory with an arbitrary gauge group.

One type of mixed anomaly merits special consideration — the (four–dimensional) "gravitational" anomaly[41] that is generated by the Feynman diagrams in Fig. 12. This anomaly may arise in a U(1) gauge theory coupled to Einstein gravity; it is proportional to

$$\mathrm{tr}\ (Q)\ , \tag{4.15}$$

where Q is the U(1) generator and the trace indicates a sum over all spin$-1/2$ Weyl fermions.

If tr $Q \neq 0$, then the effective action Γ will in general fail to be invariant under both U(1) gauge transformations and general coordinate transformations (or local Lorentz transformations). However, just as for the SU(2) × U(1) mixed anomaly discussed above, there is an appropriate choice of local counterterms such that Γ respects general covariance, but not the U(1) gauge symmetry. This choice of local counterterms is forced upon us, if we wish to ensure that the graviton remain exactly massless.

With this choice of local counterterms, the variation of the effective action under the U(1) gauge transformation eq. (2.3) is determined to be[41][*]

$$\delta_{\omega_1}\Gamma \; = \; \frac{1}{(24)(16\pi^2)} \; \mathrm{tr} \; (Q) \int d^4x \; \sqrt{g}\omega R_{\mu\nu\sigma\tau} \; \tilde{R}^{\mu\nu\sigma\tau} \quad . \tag{4.16}$$

Here $R_{\mu\nu\sigma\tau}$ is the Riemann curvature tensor and $\tilde{R}^{\mu\nu\sigma\tau} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}R_{\alpha\beta}{}^{\sigma\tau}$ in its dual. The anomalous variation of Γ under a U(1) gauge transformation may now be canceled by the local counterterm

$$\mathcal{L}_\theta \; = \; - \frac{1}{(24)16\pi^2} \; (\; \mathrm{tr} \; Q) \; \theta\sqrt{g} \; R_{\mu\nu\sigma\tau} \; \tilde{R}^{\mu\nu\sigma\tau} \quad , \tag{4.17}$$

where θ transforms under a U(1) gauge transformation as in eq. (2.5).

As in our previous analysis, we may infer from eq. (4.17) by means of a power counting analysis an inequality constraint on the mass μ of the U(1) gauge boson. The form of this constraint differs from eq. (3.1) and eq. (4.13), however, because the relevant Feynman diagrams are more sensitive than before to the ultraviolet cutoff Λ, due to the nonrenormalizability of gravity. The lower bound

$$\frac{\mu}{e} \; \gtrsim \; \frac{\kappa^2\Lambda^3}{64\pi^3} \; |\mathrm{tr} \; (Q)| \quad , \tag{4.18}$$

analogous to eq. (3.1), may be extracted from the Feynman diagrams in Fig. 13. Here $\kappa = \sqrt{16\pi G}$ is the gravitational coupling constant, where G is Newton's constant. This bound was derived by Alvarez–Gaumé and Witten[41].

As before, we can construct an underlying anomaly–free gauge theory in which the bound eq. (4.18) is saturated. In this underlying theory, the U(1) gauge symmetry undergoes the Higgs mechanism, and large masses of order Λ are acquired by a set of fermions with U(1) charges such that $\mathrm{tr} \; Q^3 = 0$ and $\mathrm{tr} \; Q \neq 0$. Eq. (4.18) may then be interpreted as an upper bound on the ultraviolet cutoff Λ of an effective field theory

---

[*] This follows from the calculation in [42] of the diagrams in Fig. 12.

from which the heavy fermions have been integrated out. Because of quantum gravity effects, the mass scale at which calculability fails in the effective field theory cannot be pushed up arbitrarily high relative to the photon mass μ. Of course, because Einstein gravity is nonrenormalizable, calculability must eventually fail in any case at a mass scale of order $4\pi/\kappa$, even if there is no anomaly. But if $\operatorname{tr} Q \neq 0$, the mass scale Λ at which new physics must be encountered may be many orders of magnitude below the scale at which gravity becomes strong.

5. The Non–Abelian Gauge Anomaly

In an Abelian gauge theory (in four dimensions) with anomalous fermion content, we derived the inequality eq. (3.1). This inequality, an upper bound on the ultraviolet cutoff Λ, identifies the mass scale at which calculability breaks down and new physics must appear. We have also seen that the inequality may be saturated in an effective field theory describing the low–energy limit of a renormalizable gauge theory that undergoes the Higgs mechanism. We noted that our construction of an effective field theory that saturates the inequality takes advantage of the fact that electric charge is not quantized in an Abelian gauge theory. In view of this observation, we may anticipate that the bound on the ultraviolet cutoff in an anomalous *non–Abelian* gauge theory will take a quite different form than in an Abelian theory, since charge *is* quantized in the non–Abelian case. The analysis of anomalous non–Abelian gauge theories will be our task in this and the next section.

Our discussion of anomalous non–Abelian gauge theories will in fact follow closely our discussion of Abelian theories. We will introduce fictitious Goldstone boson fields, and will construct a local interaction involving Goldstone bosons and gauge fields that transforms appropriately under a gauge transformation so as to cancel the gauge anomalies of the fermions. This interaction is the gauged Wess–Zumino term[8,9]. As before, the Goldstone bosons are purely a gauge artifact; they, and the Wess–Zumino term, are completely removed when we fix the unitary gauge. Introducing the Goldstone bosons, then, does not modify the physics of the anomalous gauge theory, but merely simplifies the analysis of its behavior in perturbation theory.

As in an Abelian theory, at least some gauge bosons acquire mass from radiative corrections due to the anomaly. And we may infer a lower bound on gauge boson masses (in units of the ultraviolet cutoff) by power counting of Feynman diagrams. The essential new feature that distinguishes a non–Abelian theory from an Abelian theory is that the non–Abelian structure of the gauge group induces nontrivial Goldstone boson self–couplings. The self–couplings, described by a nonrenormalizable nonlinear sigma model, have strengths that are entirely determined by gauge boson masses, and are otherwise independent of the strength of the anomaly.

Furthermore, an upper bound on the ultraviolet cutoff Λ can be derived by analyzing this nonlinear sigma model, without any further reference to the anomaly. Thus, the anomaly plays only a qualitative role in the derivation of this bound, by providing a mechanism for the generation of gauge boson masses. In contrast to the Abelian case, the strength of the anomaly does not enter into the quantitative form of the bound.

Consequently, our analysis can be just as well described as a derivation of an upper bound on the cutoff Λ in *massive* Yang–Mills theory. Massive Yang–Mills theory (like massive electrodynamics) can be expressed in a gauge–invariant form at the expense of introducing gauge artifact "Stueckelberg" fields. But in Yang–Mills theory, these Stueckelberg fields have nonrenormalizable self interactions, which is one way of understanding why massive Yang–Mills theory (unlike massive electrodynamics) is a nonrenormalizable theory. Up to group–theoretic factors, the bound on the cutoff has the form

$$\Lambda \ \lesssim \ 4\pi(\mu/e) \quad , \tag{5.1}$$

where μ is a typical gauge boson mass, e is the gauge coupling, and $F \equiv \mu/e$ is the symmetry–breaking mass scale of the associated nonlinear sigma model.

It is easy to interpret this bound in terms of an underlying anomaly–free renormalizable theory that reduces to our anomalous effective field theory in the low–energy limit. If gauge bosons acquire masses of order $\mu \sim eF$ due to the Higgs mechanism

at the characteristic mass scale F, then some fermions might also acquire masses of order F, leaving behind light fermions in an anomalous representation of the gauge group. But neither the heavy fermions nor the Higgs bosons can be expected to have masses much larger than F. It is no great surprise, therefore, to discover that the effective field theory, with heavy fermions and Higgs bosons integrated out, no longer provides an adequate description of the physics at a mass scale comparable to F.

The analysis outlined above will be described in more detail below. But before launching into that analysis, we should clarify what we mean by the statement that an anomalous theory has gauge group G. The meaning of this statement may not be apparent if the local G symmetry is spoiled by the anomaly. An essential ingredient in our analysis is that, in spite of the anomaly, it is in fact possible to quantize the anomalous theory in a manner that respects the Ward identities associated with local G invariance, at least order–by–order in perturbation theory.[★] That this is possible is clear once we recognize that the anomalous theory can arise as the low–energy limit of an underlying anomaly–free theory that undergoes the Higgs mechanism. Although the Higgs mechanism may remove any remnant of G symmetry from the mass spectrum of gauge bosons or light fermions, the effective action is nonetheless powerfully constrained by the underlying gauge invariance. (For example, there is a single renormalized gauge coupling, if G is simple). We will always choose to quantize the anomalous gauge theory so as to preserve the Ward identities; it is only because we make this choice that we are able to formally restore the local G symmetry by introducing a suitably chosen multiplet of Goldstone boson fields.

Naturally, there are many possible effective gauge theories with gauge group G and specified fermion content, just as there are many possible patterns of gauge symmetry breakdown in a G gauge theory with a sufficiently complicated Higgs sector. For example, if the theory contains fermions in an anomalous representation of G that is anomaly free with respect to a subgroup H of G, then it is possible to consistently quantize the theory while the H gauge bosons remain exactly massless. (Indeed, it

---

★ That is, BRST[10] invariance is preserved in the gauge–fixed theory. See [5] for a discussion of BRST invariance in effective field theory language.

is not even necessary for the fermions to provide complete representations of G.) For the sake of definiteness, however, we will consider in this section theories with a G gauge anomaly in which *global G* invariance is explicitly maintained. Thus, all gauge bosons are degenerate in mass, if G is simple, and the fermions provide a complete (anomalous) representation of G. The discussion will be generalized in Section 6. In particular, we will describe there the case in which the gauge bosons associated with an anomaly–free subgroup H of G remain exactly massless. That the anomaly may be manifested in various ways in a theory with given gauge group and fermion content reflects our freedom to choose the local counterterms in the effective action of the theory.

In the remainder of this paper, we will adopt different notational conventions than those established in Sections 2–4. Specifically, we will work in Euclidean space and express the gauge field as a differential form, following the notation of [40,29]. Differential form notation (which I have deliberately avoided up until now) will allow us to write compactly some algebraic expressions that would otherwise be quite cumbersome.

In this notation, A denotes the one–form

$$A \;=\; -ie \, T^a \, A^a_\mu \, dx^\mu \quad , \tag{5.2}$$

where e is the gauge coupling, T^a is a hermitian generator of G normalized so that (in the fundamental irreducible representation f of G)[†]

$$\mathrm{tr}_f(T^a T^b) \;=\; \delta^{ab} \quad , \tag{5.3}$$

and A^a_μ is a conventionally normalized gauge field. Notice that A is *anti*hermitian.

---

† This is a departure from the normalization $tr_f(T^a T^b) = \frac{1}{2}\delta^{ab}$ used in eq. (4.8)

The covariant derivative D acting on a form is defined by

$$D = d + A \quad , \tag{5.4}$$

where d is the ordinary exterior derivative. Under a finite gauge transformation $\Omega(x)\epsilon G$, A transforms according to

$$A \rightarrow A^\Omega \equiv \Omega^{-1}(d + A)\Omega \quad , \tag{5.5}$$

and under an infinitesimal gauge transformation $\Omega = \mathbf{1} + \omega$, A transforms as

$$\delta_\omega A = [D, \omega] = d\omega + [A, \omega] \quad . \tag{5.6}$$

The Yang–Mills field strength may be expressed as a curvature two–form

$$F = dA + A^2 = F_{\mu\nu} \frac{1}{2} dx^\mu dx^\nu \quad , \tag{5.7}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] \quad . \tag{5.8}$$

If G gauge fields are coupled to fermions in the representation R of G, then the fermions may be integrated out, yielding the factor $\exp(-\Gamma_R[A])$, where $\Gamma_R[A]$ is the (Euclidean) effective action. If the representation R is such that

$$\mathrm{tr}_R\left(T^a\{T^b, T^c\}\right) \neq 0 \quad , \tag{5.9}$$

there is a gauge anomaly, and Γ_R is not a gauge–invariant functional of A. The transformation property of Γ_R under the infinitesimal gauge transformation eq. (5.6) is uniquely determined if local counterterms in Γ_R are chosen so that global G invariance is maintained; it transforms as[43]

$$\delta_\omega \Gamma_R[A] = \frac{1}{24\pi^2} \int \mathrm{tr}_R\left(\omega d\left[AdA + \frac{1}{2}A^3\right]\right) \quad . \tag{5.10}$$

The right–hand side of eq. (5.10) is actually of order e^2; we absorbed a factor of e into the gauge field A in eq. (5.2).

As in Section 2, we may introduce a fictitious Goldstone boson field in order to simplify the perturbative analysis of this anomalous theory. The Goldstone boson field $g(x)$ takes values in the group G, and transforms under a finite local gauge transformation as

$$g(x) \to \Omega^{-1}(x)g(x) \quad . \tag{5.11}$$

It is evident that $g(x)$ is purely a gauge artifact; we may impose the unitary gauge condition $g(x) = \mathbf{1}$ to completely eliminate the Goldstone degrees of freedom. But in terms of g and A it is easy to construct a local interaction term that, when added to the action, restores invariance under local gauge transformations. This is the Wess–Zumino term[8,9]

$$\Gamma_R^{(\text{WZ})}[A, g] \;=\; \Gamma_R[A^g] \;-\; \Gamma_R[A] \quad . \tag{5.12}$$

Obviously, $\Gamma_R[A^g] = \Gamma_R[A] + \Gamma_R^{(\text{WZ})}[A, g]$ is invariant under the local transformation $A \to A^\Omega$, $g \to \Omega^{-1}g$. Furthermore, although $\Gamma_R[A]$ itself is of course not a local functional of A, $\Gamma_R^{(\text{WZ})}[A, g]$ *is* local in spacetime.

An explicit expression for $\Gamma_R^{(\text{WZ})}$ can be obtained by integrating eq. (5.10), which says that

$$\Gamma_R[A^{g+\delta g}] \;-\; \Gamma_R[A^g] \;=\; \frac{1}{24\pi^2} \int d^4x \, \text{tr}_R \left(g^{-1}\delta g d\left[A^g dA^g + \frac{1}{2}(A^g)^3 \right] \right) \quad , \tag{5.13}$$

if $\mathbf{1} + g^{-1}\delta g$ is an infinitesimal gauge transformation. We may calculate $\Gamma_R^{(\text{WZ})}[A, g]$ by arbitrarily choosing a function $g(x, s), s\epsilon[0, 1]$, that interpolates between $g(x, 0) = \mathbf{1}$ and $g(x, \mathbf{1}) = g(x)$. The result is[44,45,29]

$$\Gamma_R^{(\text{WZ})}[A, g] \;=\; \frac{1}{48\pi^2} \int_0^1 ds \int \text{tr}_R \left[g^{-1}\partial_s g (g^{-1}dg)^4 \right]$$

$$- \frac{1}{48\pi^2} \int \text{tr}_R \left[dgg^{-1}\left(AdA + dAA + A^3 - \frac{1}{2}Adgg^{-1}A - (dgg^{-1})^2 A \right) \right] \quad .\tag{5.14}$$

All terms on the right–hand side of eq. (5.14) are obviously local except for the first term, which requires special comment. This term is actually the integral over a five–

dimensional manifold with boundary of a five–form that is locally exact. Thus, it is invariant under small deformations of the interpolating function $g(x, s)$, and it can be regarded as an effectively local functional on spacetime[9]. There may, in general, be different interpolations $g(x, s)$ that cannot be smoothly deformed one to another, such that a change in the interpolation modifies $\Gamma_R^{(\mathrm{WZ})}$ by an integer multiple of $2\pi i$. But this ambiguity has no effect on $\exp(-\Gamma_R^{(\mathrm{WZ})})$.

As in the Abelian case, we must allow the gauge fields to have explicit masses in an anomalous gauge theory. If global G invariance is to be maintained, and G is simple, then all gauge bosons have equal masses. Making use of the Goldstone boson field $g(x)$, we may construct a unique local mass term that is quadratic in A and has local G invariance; it is

$$
\begin{aligned}
\mathcal{L}_{\mathrm{mass}} &= -\frac{1}{2}(\mu^2/e^2)\mathrm{tr}_f\left[(A^g)^\mu (A^g)_\mu\right] \\
&= -\frac{1}{2}(\mu^2/e^2)\mathrm{tr}_f\left[(\partial_\mu g)g^{-1} + A_\mu\right]^2 \; .
\end{aligned}
\tag{5.15}
$$

This is the "Stueckelberg" form of the mass term, and it reduces to the standard form in the unitary gauge $g = \mathbf{1}$.

To complete the construction of the anomalous gauge theory, we must specify how path integral quantization is to be performed. In the unitary gauge theory with $g = \mathbf{1}$, and with nonvanishing gauge boson masses, we may integrate over all gauge field histories; there is no need for further gauge fixing. Alternatively, we obtain nicer ultraviolet behavior in perturbation theory if we choose a different gauge slice such that $g(x)$ is not eliminated. The integration over this slice may be performed by the Faddeev–Popov method, but first a measure must be defined for the integral over $g(x)$. This measure should be local and should respect the local G invariance; hence it is natural to choose the invariant Haar measure

$$
d\mu_{\mathrm{Haar}}(g) = \prod_x \prod_{a=1}^n (g^{-1}dg)^a \; ,
\tag{5.16}
$$

where n is the dimension of the gauge group G.[*]

Of course, implicit in the definition of the path integral measure is a short–distance regulator that controls ultraviolet divergences. Since our analysis will be completely perturbative, we need not address the problem of constructing a nonperturbative regulator. Nor need we describe explicitly our method of regulating Feynman diagrams. It will be understood that our regulator cuts off the loop momenta of Feynman diagrams at a mass scale of order Λ, and that local counterterms are to be introduced as needed in order to ensure that the BRST[10] identities are satisfied order–by–order in perturbation theory.

Now integrating g and A over a gauge slice evidently agrees with the procedure of integrating over all A in the unitary gauge, since the gauge choice $g(x) = \mathbf{1}$ may be implemented (at least perturbatively) without any need to introduce ghosts. For the purpose of perturbative power counting, we prefer to use the 't Hooft gauges. In order to write out the necessary gauge fixing term, it is convenient to introduce the notation

$$g(x) \;=\; \exp\big[\phi(x)/F\big] \quad, \tag{5.17}$$

where $\phi = -iT^a\phi^a$ is an antihermitian scalar field that takes values in the Lie algebra of G, and F has the dimensions of mass. In terms of ϕ, the gauge field mass term eq. (5.15) may be expanded as

$$\mathcal{L}_{\text{mass}} \;=\; -\frac{1}{2}(\mu^2/e^2)\text{tr}_f\left[\frac{1}{F}\,\partial_\mu\phi + A_\mu + \dots \,\right]^2 \tag{5.18}$$

and the measure eq. (5.16) becomes

$$d\mu(\phi) \;=\; \prod_x e^{h(\phi)}\, d\phi^1\dots d\phi^n \quad. \tag{5.19}$$

(Here $h(\phi)$ may be regarded as an ultra–local nonpolynomial self–interaction for ϕ that must be included in order to maintain local G invariance.) The mixing between

---

[*] The most general measure with the desired properties has the form $d\mu(g) = d\mu_{\text{Haar}}(g)exp(-\delta S)$, where δS is a local gauge-invariant counterterm that can be absorbed into the action.

ϕ and A in eq. (5.18) is eliminated by the gauge fixing term

$$
\mathcal{L}_{\text{gauge fixing}} \;=\; -\frac{1}{2\alpha}\mathrm{tr}_f\left[\frac{1}{e}\partial^\mu A_\mu + \alpha\mu\phi\right]^2 \quad , \tag{5.20}
$$

where α is an arbitrary gauge parameter. We have chosen $F = \mu/e$, so that $i\phi^a$ is a conventionally normalized real scalar field.

To complete the 't Hooft gauge formulation of the theory, we must include the Faddeev–Popov ghosts. The ghost term in the effective action can be systematically expanded in powers of ϕ. Since

$$
\delta_\omega(\partial^\mu A_\mu + e\alpha\mu\phi)
$$
$$
= \; \partial^\mu[D_\mu, \omega] - e\alpha\mu F\omega + 0(\phi) \quad , \tag{5.21}
$$

the ghost term is

$$
\mathcal{L}_{\text{ghost}} = \mathrm{tr}_f\eta^\dagger[-\partial^\mu[D_\mu, \eta] + \alpha\mu^2\eta] + \dots \tag{5.22}
$$

where $\eta = -iT^a\eta^a$ is a Grassmann variable.

Notice that the (Euclidean) gauge field propagator

$$
\Delta^{ab}_{\mu\nu} \;=\; \frac{\delta^{ab}}{k^2 + \mu^2}\left[\delta_{\mu\nu} + (\alpha - 1)\frac{k^\mu k^\nu}{k^2 + \alpha\mu^2}\right] \tag{5.23}
$$

has a physical pole at $k^2 = -\mu^2$ and a gauge–artifact pole at $k^2 = -\alpha\mu^2$. The ϕ and η propagators also have poles at $k^2 = -\alpha\mu^2$, and the singularities at $k^2 = -\alpha\mu^2$ always cancel in gauge–invariant quantities. For many purposes, the most convenient gauge choice is the 't Hooft–Landau gauge $\alpha = 0$. In this gauge the A propagator is purely transverse, ϕ and η are massless, and the ϕ field is derivatively coupled. Unitary gauge is recovered from the 't Hooft gauge in the limit $\alpha \to \infty$, for in this limit ϕ and η become infinitely massive and decouple. Since unphysical degrees of freedom cease to propagate as $\alpha \to \infty$, unitarity is manifestly satisfied in this gauge.

Now we consider the upper limit on the ultraviolet cutoff Λ of the anomalous gauge theory. As we have already discussed, this bound is actually insensitive to the strength of the gauge anomaly. It is really a bound on Λ/μ in massive Yang–Mills theory, and is most conveniently derived in terms of the sigma–model Lagrangian eq. (5.15). Because the sigma–model is nonrenormalizable, we expect that there is a nonperturbative bound on the renormalized value of $F = \mu/e$, in units of the cutoff Λ. Although deriving this nonperturbative bound is beyond our analytical ability, we may estimate the maximal value of F as that value for which perturbation theory in F^{-1} breaks down. The breakdown of perturbation theory signifies a failure of quasi–locality and of unitarity. It occurs at a mass scale where new physical degrees of freedom must be introduced in order to restore a local description of the physics.

We may ignore the gauge fields and study the sigma model

$$\mathcal{L} = -\frac{1}{2}F^2 \text{tr}_f(\partial^\mu g \partial_\mu g^{-1}) + \dots \quad , \tag{5.24}$$

where the ellipsis denotes terms with more derivatives that become irrelevant in the low–energy limit. Feynman rules for the sigma model may be extracted from the path integral by expanding $g = \exp(\phi/F)$ in powers of ϕ. One finds, for example, that the most ultraviolet divergent part of the one–loop diagram of Fig. 14 induces a renormalization of F^2 of order

$$\left(\frac{\delta F^2}{F^2}\right)_{1-\text{loop}} \sim \frac{C_G}{16\pi^2}\frac{\Lambda^2}{F^2} \quad , \tag{5.25}$$

where C_G is the quadratic Casimir invariant of the adjoint representation of G. Similarly, the leading renormalization of F^2 in L–loop order has the form

$$\left(\frac{\delta F^2}{F^2}\right)_{L-\text{loop}} \sim \left(\frac{C_G}{16\pi^2}\frac{\Lambda^2}{F^2}\right)^L \quad . \tag{5.26}$$

Thus

$$\Lambda \lesssim 4\pi F C_G^{-1/2} \tag{5.27}$$

is the condition that must be satisfied for renormalization effects to be small; the

right–hand side of eq. (5.27) is the mass scale at which the sigma model becomes strongly coupled.

Indeed, eq. (5.27) may also be interpreted as the condition for calculability. Divergent loop diagrams also induce local interactions with more derivatives, schematically of the form

$$\mathcal{L}_{2n} \sim F^2\left[\lambda^{(2n)} \text{ tr } (g^{-1}\partial g)^{2n}\right] \quad . \tag{5.28}$$

The leading L–loop renormalization of $\lambda^{(2n)}$ is of order

$$(\delta\lambda^{(2n)})_{\text{L-loop}} \sim \Lambda^{2-2n}\left(\frac{C_G}{16\pi^2}\frac{\Lambda^2}{F^2}\right)^L \quad , \tag{5.29}$$

for $L \geq n-1$. Thus, when the expansion parameter $C_G\Lambda^2/16\pi^2 F^2$ is of order one, an infinite number of counterterms give sizable contributions to processes at the energy scale Λ, and calculability is lost. Therefore, eq. (5.27), or

$$\mu \gtrsim C_G^{1/2}(e\Lambda/4\pi) \tag{5.30}$$

is the analog of eq. (3.1) that applies in an anomalous non–Abelian gauge theory (that is *globally* gauge invariant.)

Our discussion of the Abelian case has prepared us to expect that the bound eq. (5.27) can be understood in terms of an anomaly–free renormalizable gauge theory that undergoes the Higgs phenomenon at the symmetry breaking mass scale F. To grasp this interpretation of eq. (5.27), we should consider an underlying theory in which global G symmetry survives the Higgs mechanism, since global G invariance was assumed in the derivation of eq. (5.27). The desired symmetry–breaking pattern can be achieved if the underlying gauge theory respects the symmetry $G_{\text{gauge}} \times G_{\text{global}}$, which is spontaneously broken to the "diagonal" subgroup.

Suppose, for example, that the gauge group is $G = \text{SU(N)}$. A model with $\text{SU(N)}_{\text{gauge}} \times \text{SU(N)}_{\text{global}}$ symmetry can be constructed that contains N Higgs fields,

each transforming as the fundamental representation N of SU(N)$_{\text{gauge}}$. These N Higgs fields may be assembled into an $N \times N$ matrix Φ that transforms according to

$$\Phi \rightarrow \Omega \Phi U^\dagger \quad , \tag{5.31}$$

where $\Omega \in \text{SU(N)}_{\text{gauge}}$ and $U \in \text{SU(N)}_{\text{global}}$. The most general quartic potential for Φ that is invariant under SU(N)$_{\text{gauge}} \times$ SU(N)$_{\text{global}}$ is

$$
\begin{aligned}
V(\Phi) \;=\; & \lambda_1 \Big[\text{ tr } (\Phi^\dagger \Phi - v^2 \mathbf{1}) \Big]^2 \\
& + \; \lambda_2 \text{ tr } (\Phi^\dagger \Phi - v^2 \mathbf{1})^2 \quad .
\end{aligned}
\tag{5.32}
$$

Provided that $v^2 > 0$, $\lambda_2 \geq 0$ and $\lambda_2 + \lambda_1 N \geq 0$, the minimum of this potential occurs at*

$$\Phi \;=\; \Phi_0 \;\equiv\; v\mathbf{1} \quad , \tag{5.33}$$

up to an SU(N)$_{\text{gauge}} \times$ SU(N)$_{\text{global}}$ rotation. This expectation value breaks the SU(N)$_{\text{gauge}}$ symmetry completely, but leaves intact the SU(N)$'_{\text{global}}$ symmetry generated by $T^a_{\text{gauge}} + T^a_{\text{global}}$. Through the Higgs phenomenon, the gauge bosons acquire the mass matrix

$$
\begin{aligned}
\mu^2_{ab} \;=\; & e^2 \left(T^a_{\text{gauge}} \Phi_0, T^b_{\text{gauge}} \Phi_0 \right) \\
\;=\; & e^2 v^2 \text{ tr } (T^a T^b) \;=\; e^2 v^2 \delta_{ab} \quad .
\end{aligned}
\tag{5.34}
$$

Thus, v is the symmetry breaking scale F, in our earlier notation.

Of the $2N^2$ real scalar fields that comprise Φ, $N^2 - 1$ are eaten by the gauge bosons. Of those remaining, $N^2 - 1$ transform as the adjoint representation of SU(N)$'_{\text{global}}$,

---

$\star$ Minimization of this potential is discussed in, for example, [46].

and have mass

$$m_{\text{Adj}}^2 = 2\lambda_2 v^2 \quad , \tag{5.35}$$

to leading order in λ. There is also a $\text{SU(N)}'_{\text{global}}$ singlet with mass

$$m_{\text{Sing}}^2 = 2(\lambda_1 N + \lambda_2)v^2 \quad , \tag{5.36}$$

to leading order, and a massless singlet. The massless singlet is the Goldstone boson associated with the spontaneously broken U(1) global symmetry $\Phi \rightarrow e^{i\alpha}\Phi$ of the potential eq. (5.32).

In terms of this underlying theory, we can understand the origin of the group–theoretic factor $C_G = N$ in eq. (5.27). In the underlying theory there is an upper bound on the Higgs mass m_{Adj}, in units of the gauge boson mass $\mu = eF$. This bound arises because scalar self–couplings are nonasymptotically free, and hence (presumably) trivial in the continuum limit[3,47]. The underlying theory itself therefore requires a finite ultraviolet cutoff and a strong renormalized coupling can be attained at a renormalization mass scale M only if the cutoff Λ is not far above M. In leading order in perturbation theory, the mass ratio is

$$\frac{m_{\text{Adj}}^2}{\mu^2} = \frac{2\lambda_2}{e^2} \quad , \tag{5.37}$$

where, roughly speaking, the couplings are renormalized at the scale F. To increase the mass ratio, we make the scalar self–coupling stronger, but this is achieved at the cost of lowering the cutoff Λ. Since we must require that the cutoff exceed the mass m_{Adj}, there is a fundamental limitation on how large the mass ratio can become.

This upper bound is intrinsically nonperturbative, and it is beyond our ability to compute it precisely.[†] But we can estimate it by invoking the criterion that perturbation theory breaks down when λ is of order its maximum allowed value. Now we note

---

† A nonperturbative method for computing the upper bound on the Higgs mass, in a related context, was sketched in [48].

that the quartic interaction in eq. (5.32) couples each Higgs "flavor" to N different species. Hence the perturbative expansion parameter is actually $N\lambda_2/16\pi^2$, and the Higgs sector becomes strongly coupled for

$$N \frac{\lambda_2}{16\pi^2} \sim 1 \quad . \tag{5.38}$$

We therefore estimate that

$$\frac{m_{\text{Adj}}}{\mu} \lesssim \frac{4\pi}{e\sqrt{N}} \quad . \tag{5.39}$$

This is in accord with eq. (5.27), if we interpret the Higgs mass m_{Adj} as the cutoff of the low–energy effective field theory. Like eq. (5.27), eq. (5.39) is derived without any regard for the fermion content of the theory. Rather, it is a bound on the mass scale of the physics responsible for driving the Higgs mechanism.

(The theory in which eq. (5.39) was derived also contains the massless U(1) Goldstone boson, which did not appear in the theory in which eq. (5.27) was derived. But this extra Goldstone boson had no effect on the discussion. It decouples from low–energy physics, since its derivative couplings are suppressed by powers of $(Nv^2)^{-1}$. If we wish, we may imagine giving the unwanted Goldstone boson a large mass of order $\sqrt{N}v$, by, for example, gauging the spontaneously broken U(1) symmetry.)

If we do include chiral fermions in the underlying gauge theory, then some fermions may acquire large masses due to the Higgs mechanism, and we may contemplate integrating out the heavy fermions, thereby constructing an effective field theory with anomalous light fermion content. (If global G invariance is maintained, then all fermions in an irreducible representation of G acquire equal masses, so it is guaranteed that the light fermions comprise a complete representation of G.) As in the Abelian case discussed in Section 3, we may rotate the Goldstone boson fields out of the heavy fermion mass terms before the heavy fermions are integrated out. The Wess–Zumino term eq. (5.14) arises as the nontrivial Jacobian of this chiral rotation[44,45,27]. The gauge noninvariance of the Wess–Zumino term compensates for the anomalous light fermion content so as to ensure that the low–energy effective theory is gauge invariant.

Let us now return to the discussion of mixed gauge anomalies that we began in Section 4. We observed there that if a theory with gauge group $G \times$ U(1) has a $G - G - $U(1) mixed anomaly, then by an appropriate choice of local counterterms it is possible to preserve the invariance of the fermionic effective action under *either* G or U(1) gauge transformations, but not *both*. In our discussion in Section 4, we chose to preserve the local G symmetry, and then noted that local U(1) symmetry could also be restored by introducing an appropriate Wess–Zumino term. The required Goldstone boson transformed under U(1) but not under G. We will now describe an alternative procedure, in which the Goldstone boson fields transform under G but not under U(1).

We anticipate that this procedure is feasible, because we can imagine construct- ing an underlying anomaly–free theory with $G \times$ U(1) gauge symmetry such that G but not U(1) is spontaneously broken. Then fermions with nontrivial $G - G - $U(1) anomalies may acquire large masses and can be integrated out. The resulting low– energy effective field theory has fermion content with a mixed anomaly, but the U(1) "photon" remains exactly massless.

To be concrete, consider an SU(2) $\times$ U(1) gauge theory, as in Section 4. By including in the effective action Γ the counterterm eq. (4.8) with an appropriate coefficient, the variation of Γ under a gauge transformation becomes eq. (4.5)–(4.6), with $C_1 = 0$ and $C_2 = 1$. In our Euclidean differential form notation, this is

$$
\begin{aligned}
\delta_{\omega_1} \Gamma &= 0 \quad, \\
\delta_{\omega_2} \Gamma &= \frac{1}{8\pi^2} \int \mathrm{tr}_R[\omega_2 F_1 dA_2] \quad.
\end{aligned}
\tag{5.40}
$$

(The subscripts 1 and 2 distinguish U(1) and SU(2) gauge fields.) If we now introduce a Goldstone boson field $g(x)$ that takes values in SU(2), then we may construct in terms of g a Wess–Zumino term that cancels the anomalous variation of Γ under an SU(2) gauge transformation. The derivation of the Wess–Zumino term follows the steps of eq. (5.11)–(5.14) above; it is formally similar to the derivation of the SU(2)

Wess–Zumino term in two spacetime dimensions. The result is

$$\Gamma_R^{(\mathrm{WZ})}[A_1, A_2, g] = - \frac{1}{24\pi^2} \int\limits_{M^5} \mathrm{tr}_R\Big[F_1(dg g^{-1})^3\Big]$$

$$- \frac{1}{8\pi^2} \int\limits_{M^4} \mathrm{tr}_R[F_1 dg g^{-1} A_2] \quad . \tag{5.41}$$

The first term on the right–hand side of eq. (5.41) is the integral of a locally exact five–form over a five–dimensional manifold M^5 whose boundary is Euclidean spacetime, and the second term is integrated over four–dimensional Euclidean spacetime M^4. The first term can be expanded in powers of ϕ, where $g = \exp(\phi/F)$, as an infinite series of local interaction terms.

Once we have introduced the Goldstone boson field g, we may proceed to derive an upper bound on the ultraviolet cutoff Λ, as described earlier in this section. We obtain, again,

$$\Lambda \lesssim 4\pi(\mu/e_2) \quad , \tag{5.42}$$

where μ is the mass of the SU(2) gauge bosons, e_2 is the SU(2) gauge coupling, and global SU(2) invariance has been assumed.

It is also easy to construct an underlying anomaly–free gauge theory such that the bound eq. (5.42) is saturated. Consider an SU(2) × U(1) gauge theory that contains a Higgs doublet ϕ with vanishing U(1) charge. Among the (Weyl) fermions in the theory are an SU(2) doublet ψ_D with U(1) charge Q and SU(2) singlets ψ_S, ψ'_S, each with U(1) charge $-Q$. Thus there is an SU(2) × U(1)–invariant Yukawa coupling

$$\mathcal{L}_{\mathrm{Yuk}} = -\lambda\, \psi_S \psi_D \phi - \lambda' \psi'_S \psi_D \phi^c \quad . \tag{5.43}$$

The set of fermions ψ_D, ψ_S, ψ'_S has a vanishing U(1)^3 anomaly, but a nonvanishing mixed anomaly,

$$\mathrm{tr}\,(T^a T^b Q) = Q\delta^{ab} \quad . \tag{5.44}$$

We may imagine that this mixed anomaly is canceled by additional fermions that are not Yukawa coupled to ϕ. (ψ_D also contributes a nonperturbative SU(2) anomaly[49]

that must be canceled; or we can remove the SU(2) anomaly by merely doubling the fermion content.)

When ϕ acquires a vacuum expectation value, the SU(2) gauge symmetry (but not the U(1) gauge symmetry) is spontaneously broken. The fermions ψ_D, ψ_S, ψ'_S all acquired large masses, if the Yukawa couplings are strong. If we integrate out the heavy fermions and the Higgs boson, then the effective field theory that we obtain contains remaining light fermions with a nontrivial mixed anomaly; yet the photon is exactly massless. The bound eq. (5.42) may be saturated, where Λ is the scale of the masses of the heavy fermions and Higgs boson.

(Actually, we derived eq. (5.42) under the assumption of unbroken SU(2) global symmetry, and the model that we have just described does not respect this symmetry. However, the model does have a "custodial" SU(2) global symmetry[50] for $\lambda = \lambda'$ and in the limit $e_2 \to 0$. Therefore eq. (5.42) does apply to our model, even if the Yukawa couplings are strong, provided that $\lambda = \lambda'$ and we work perturbatively in the gauge coupling e_2.)

To conclude this section, I wish to comment on a possible objection to the power–counting analysis that I have described here. It has been suggested[51,25] that the apparent nonrenormalizability of the nonlinear sigma model in four–dimensional space-time is actually an illusion, for the ultraviolet behavior of perturbation theory can be greatly improved by a clever resummation. Advocates of this view emphasize that the action of the sigma model contains terms that are quadratic in the field ϕ and contain four or more derivatives. If such higher–derivative terms are included in the perturbative inverse propagator, then the propagator acquires the momentum dependence

$$\Delta(p^2) \sim (p^2 + \frac{a}{\Lambda^2} p^4 + \ldots)^{-1} \quad . \tag{5.45}$$

Here Λ is the ultraviolet cutoff, and a is the (dimensionless) coupling constant associated with the four–derivative term in the action. Now one notes that for $|a| \gg 1$, $\Delta(p^2)$ has much milder behavior for $p^2 \sim \Lambda^2$ than the usual perturbative propagator

$\Delta \sim p^{-2}$. This improved behavior suppresses ultraviolet divergences, so that only a finite number of divergent counterterms are generated by loop corrections. Thus, it seems, by the simple trick of choosing $|a| >> 1$, the sigma model is rendered renormalizable. The cutoff Λ can be pushed up arbitrarily high without any loss of calculability (contrary to my claims in this paper).

In fact, however, this improved ultraviolet behavior comes at a heavy price. Because $\Delta(p^2)$ in eq. (5.45) has a pole at $p^2 \sim -\Lambda^2/a << \Lambda^2$, a new light excitation has appeared in the theory. Furthermore, depending on the sign of a, this new excitation is either a tachyon or a negative metric ghost. Thus, the resummed perturbation theory achieves renormalizability by power counting only at the cost of violating unitarity order–by–order.

From our point of view this is no surprise, for we regard the breakdown of perturbation theory at short distances in the nonlinear sigma model as a signal of the breakdown of quasilocality and hence of unitarity. I am not aware of any persuasive evidence in favor of the existence of a renormalizable resummation of sigma model perturbation theory (in four dimensions) that makes physical sense.[⋆]

6. Gauged G/H Models

We described in the previous section how an anomalous gauge theory with non–Abelian gauge group G can be quantized. In our construction, global G symmetry was maintained; hence, for example, all of the G gauge bosons had equal (and nonzero) masses. We have also emphasized that the anomalous gauge theory may be regarded as an effective field theory that describes the low–energy limit of an underlying anomaly–free gauge theory that has undergone the Higgs mechanism. From this point of view, it is clear that our globally G–invariant theory cannot be the most general possible anomalous G gauge theory. For we may contemplate an underlying

---

⋆ An *ad hoc* prescription has been proposed[52] that purports to restore unitarity in a theory that contains ghosts, at least order–by–order in perturbation theory. I doubt that this prescription really makes sense nonperturbatively, but I must admit that the question cannot be regarded as definitively settled. See [53,54] for relatively recent discussions.

gauge theory in which G is spontaneously broken to H. In this theory, the H gauge bosons stay massless, and the mass–matrix of the remaining massive gauge bosons is required only to be H–invariant, not G–invariant.

In this section, we will generalize the discussion of Section 5. Specifically, we will formulate a more general procedure for quantizing a G gauge theory with fermions in an anomalous representation of G, such that the gauge boson associated with a subgroup H of G remain massless, and the other gauge bosons have H–invariant masses. This procedure can be carried out if and only if the fermions are in a representation of G that is anomaly–free when construed as a representation of the subgroup H.

Following the strategy of Section 5, we will introduce gauge artifact Goldstone bosons that take values in the coset space G/H. Then we will see that a local Wess–Zumino counterterm involving gauge fields and Goldstone bosons can be constructed, with an anomalous transformation property under local G transformations that cancels the anomaly of the fermionic effective action. It is thus possible to perform path integral quantization of the theory in an 't Hooft gauge, where power counting of Feynman diagrams is easy. As before, the theory that we construct is a nonrenormalizable theory with a maximal ultraviolet cutoff Λ of order $4\pi F$, where F is a typical symmetry breaking scale associated with the Goldstone bosons.

Since our construction involves Goldstone bosons that take values in G/H, we will begin the discussion by reviewing the formalism, developed in [55], for dealing with G/H nonlinear sigma models. If a field theory is invariant under a (global) symmetry group G that is spontaneously broken into $H \subset G$, then the elements of the coset space G/H characterize the possible ways of orienting the vacuum state of the theory. We may parametrize the coset space as

$$G/H \;=\; \left\{ \Phi \;=\; g\Phi_0 \;,\; g \in G \right\} \;,\qquad (6.1)$$

where Φ_0 is an order parameter that is invariant under the subgroup H,

$$h\Phi_0 \;=\; \Phi_0 \;,\; h \in H \quad. \qquad (6.2)$$

A history for the Goldstone bosons of the theory is thus expressed as a mapping from spacetime to G/H. In view of eq. (6.1), a field that takes values in G/H may be parametrized by a G–valued function $g(x)$, but this parametrization is evidently not unique. Under the local transformation

$$h \; : \; g(x) \;\; \rightarrow \;\; g(x)h(x) \; , \;\; h(x) \in H \quad , \tag{6.3}$$

$g(x)$ is replaced by a different parametrization of the same G/H–valued Goldstone boson field. We may find it convenient, in an effective field theory for the G/H Goldstone bosons, to use this redundant parametrization of Goldstone bosons by G– valued functions. But if we use it, then the action of the theory must respect the local symmetry eq. (6.3). Furthermore, when we quantize the theory, we must "fix the gauge" by choosing a *unique* G–valued field to represent each G/H–valued field.

A particularly apt way to "fix the gauge" was advocated in [55] To describe this choice, we first note that the Lie algebra $\mathcal{G}$ admits the decomposition

$$\mathcal{G} \;\; = \;\; \mathcal{H} \; \oplus \; \mathcal{K} \tag{6.4}$$

where $\mathcal{H}$ is the Lie algebra of the unbroken group $\mathcal{H}$, and $\mathcal{K}$ is the linear space of "broken" generators (those generators of $\mathcal{G}$ that are orthogonal to $\mathcal{H}$). It is possible to choose a basis for $\mathcal{H}$

$$T^\alpha \; , \;\; \alpha \;\; = \;\; 1, 2, \ldots \quad , \tag{6.5}$$

and a basis for $\mathcal{K}$

$$X^a \; , \;\; a \;\; = \;\; 1, 2, \ldots \quad , \tag{6.6}$$

such that

$$\mathrm{tr} \, (T^\alpha T^\beta) \;\; = \;\; \delta^{\alpha\beta} \; , \;\; \mathrm{tr} \, (X^a X^b) \;\; = \;\; \delta^{ab} \; , \;\; \mathrm{tr} \, (T^\alpha X^a) \;\; = \;\; 0 \quad . \tag{6.7}$$

Now following [55], we may choose a standard element of g to represent each coset in

G/H according to

$$g \;=\; \exp(\phi) \;=\; \exp(\phi^a X^a) \quad , \tag{6.8}$$

where ϕ takes values in $\mathcal{K}$. This parametrization of G/H is locally well–defined if G is any finite–dimensional Lie group. (That is, there is a one–to–one correspondence between the elements of an infinitesimal neighborhood of a point in G/H and the elements of an infinitesimal neighborhood of the corresponding point in $\mathcal{K}$.)

The group G acts on the coset space G/H as

$$\Omega \in G \;:\; g\Phi_0 \to \Omega^{-1} g \Phi_0 \quad . \tag{6.9}$$

But left multiplication of g by Ω^{-1} does not, in general, preserve the standard parametrization of cosets in eq. (6.8). Thus, the action of G on the field ϕ has a form given by

$$\Omega \;:\; g \;\to\; \Omega^{-1} g h \quad . \tag{6.10}$$

Here $h = h(\Omega, g)$ has an implicit dependence on both Ω and ϕ; it is the "compensating" element of H that rotates $\Omega^{-1} g$ to the corresponding standard coset representative. For $\Omega \in H$, this compensator is simply $h = \Omega$, and ϕ transforms as a linear representation of H. But for $\Omega \notin H$, $h(\Omega, g)$ is a complicated nonlinear function. Thus, the action of G on ϕ implicitly indicated by eq. (6.10) defines not a linear representation, but a nonlinear "realization" of G. This realization associates with each $\Omega \in G$ a function f_Ω from G/H to G/H such that composition of functions respects the group multiplication law, or

$$f_\Omega \;\circ\; f_{\Omega'} \;=\; f_{\Omega\Omega'} \quad . \tag{6.11}$$

Now we wish to construct field theories involving G/H Goldstone bosons that are invariant under the nonlinearly realized global G symmetry. For this purpose,

consider the one–form $g^{-1}dg$ that takes values in the Lie algebra $\mathcal{G}$. Like any element of $\mathcal{G}$, $g^{-1}dg$ may be decomposed into broken and unbroken generators as

$$g^{-1}dg \;=\; (g^{-1}dg)_{\mathcal{H}} \;+\; (g^{-1}dg)_{\mathcal{K}} \quad, \tag{6.12}$$

where the subscripts $\mathcal{H}, \mathcal{K}$ denote projections onto $\mathcal{H}, \mathcal{K}$ respectively. Under the global G transformation eq. (6.10), this one–form transforms as

$$\Omega \;:\; g^{-1}dg \;\rightarrow\; h^{-1}(g^{-1}dg)h \;+\; h^{-1}dh \quad. \tag{6.13}$$

(Recall that h depends implicitly on Ω and g.) Because the adjoint action of H on $\mathcal{G}$ preserves the decomposition into $\mathcal{H}$ and $\mathcal{K}$, eq. (6.13) implies that the projected one–forms transform as

$$\Omega \;:\; (g^{-1}dg)_{\mathcal{H}} \;\rightarrow\; h^{-1}(g^{-1}dg)_{\mathcal{H}}h + h^{-1}dh \quad,$$
$$\Omega \;:\; (g^{-1}dg)_{\mathcal{K}} \;\rightarrow\; h^{-1}(g^{-1}dg)_{\mathcal{K}}h \quad. \tag{6.14}$$

We see that $(g^{-1}dg)_{\mathcal{K}}$ transforms homogeneously; it may be regarded as a "covariant" derivative of the field ϕ, and can be employed to construct invariant derivative self–interactions for ϕ. In particular, a kinetic term can be constructed for the Goldstone boson field ϕ that is quadratic in covariant derivatives; it is*

$$\begin{aligned}
\mathcal{L}(\phi) \;&=\; -\,\frac{1}{2}\,F^2\,\mathrm{tr}\,\Big[(g^{-1}dg)_{\mathcal{K}} \,*\, (g^{-1}dg)_{\mathcal{K}}\Big] \\
&=\; -\,\frac{1}{2}\,F^2\,\mathrm{tr}\,\Big[(g^{-1}\partial_\mu g)_{\mathcal{K}}(g^{-1}\partial^\mu g)_{\mathcal{K}}\Big] \quad.
\end{aligned} \tag{6.15}$$

(Here $* \, \omega$ denotes the Hodge dual of the form ω.)

$\star$ This is not the unique term with two derivatives unless $\mathcal{K}$ transforms irreducibly under H. See below.

The one–form $(g^{-1}dg)_{\mathcal{H}}$ is seen in eq. (6.14) to transform like an $\mathcal{H}$–valued connection, or gauge field. It can be used to construct covariant derivatives of "matter" fields that transform as linear representations of $\mathcal{H}$, and hence to construct G–invariant couplings of such fields to the Goldstone bosons. Specifically, if a matter field ψ transforms as a linear representation R of H, then a nonlinearly realized action of G on ψ may be defined by

$$\Omega \ : \ \psi \ \rightarrow \ D_R(h^{-1})\psi \quad , \tag{6.16}$$

where $h(\Omega, g) \in H$ is the compensating H rotation in eq. (6.10). The covariant derivative

$$D\psi \ = \ \left\{ d \ + \ D_R\left[(g^{-1}dg)_{\mathcal{H}}\right] \right\}\psi \tag{6.17}$$

then transforms homogeneously under a global G transformation,

$$\Omega \ : \ D\psi \ \rightarrow \ D_R(h^{-1})D\psi \quad . \tag{6.18}$$

Thus, the minimal coupling prescription may be invoked to construct a G–invariant coupling of ψ to the Goldstone bosons. If, for example, ψ is a left–handed Weyl fermion, then there is a G–invariant kinetic term for ψ,

$$\begin{aligned}
\mathcal{L}(\psi) \ &= \ \psi^\dagger \gamma^\mu D_\mu \psi \\
&= \ \psi^\dagger \gamma^\mu \left\{ \partial_\mu \ + \ D_R[(g^{-1}\partial_\mu g)_{\mathcal{H}}] \right\}\psi \quad .
\end{aligned} \tag{6.19}$$

Note that we need not require the fermions to provide a complete linear representation of G; it suffices that they transform as a linear representation of H.

(The coupling of fermions to Goldstone bosons in eq. (6.19) respects the global G symmetry at the classical level, but the symmetry *may* be intrinsically broken by quantum effects[56] if the fermions are in an anomalous representation of H (a representation R such that $\mathrm{tr}_R(T^\alpha\{T^\beta, T^\gamma\}) \neq 0$).) As explained in [57-59], a suitable local counterterm that restores the global G symmetry can always be found if there exists *some* linear representation of G whose H anomalies "match" those of R.)

We have seen how to construct an action that is invariant under the nonlinearly realized G symmetry. But in order to perform path integral quantization of the theory, we must also define a functional measure. This measure must be G invariant if the quantum theory is to satisfy the G Ward identities. The G–invariant measure on G/H can be expressed as

$$d\mu(\phi) = \prod_a (g^{-1}dg)^a$$
$$\equiv \exp\left[-\mathcal{L}_{\text{measure}}(\phi)\right] \prod_a d\phi^a \quad, \tag{6.20}$$

where $(g^{-1}dg)_{\mathcal{K}} = (g^{-1}dg)^a X^a$. That this is an invariant measure follows from the homogeneous transformation property eq. (6.14) for $(g^{-1}dg)_{\mathcal{K}}$, and the fact that the adjoint action of H in $\mathcal{K}$ is merely an orthogonal "rotation of axes." The invariant measure is unique up to a multiplicative rescaling. The G–invariant path integral measure, if local, is unique up to a multiplicative local G–invariant functional that can be absorbed into the (Euclidean) action of the theory. Note that if we choose to evaluate the path integral using the naive measure $\prod_a d\phi^a$ (as we do in chiral perturbation theory) then the noninvariant term $\int \mathcal{L}_{\text{measure}}(\phi)$ must be included in the action to compensate for the noninvariance of the naive measure.[60,61]

So far, we have considered G/H sigma models with global G symmetry. We can readily modify these models so that G (or a subgroup of G) is gauged; we need only construct derivatives that are *gauge* covariant. Under the *local* G symmetry, Goldstone bosons transform as

$$\Omega : g(x) \rightarrow \Omega^{-1}(x)g(x)h\big[\Omega(x), g(x)\big] \quad, \tag{6.21}$$

and we introduce a $\mathcal{G}$–valued gauge field A that transforms as

$$\Omega : A \rightarrow A^{\Omega} \equiv \Omega^{-1}(d + A)\Omega \quad. \tag{6.22}$$

Since A^g transforms according to

$$\Omega : A^g \rightarrow (A^g)^h = h^{-1}(A^g)h + h^{-1}dh \quad, \tag{6.23}$$

we observe that we may invoke the decomposition $\mathcal{G} = \mathcal{H} \oplus \mathcal{K}$ and infer the transformation properties

$$
\begin{aligned}
\Omega &: (A^g)_{\mathcal{H}} \rightarrow h^{-1}(A^g)_{\mathcal{H}}h + h^{-1}dh \quad, \\
\Omega &: (A^g)_{\mathcal{K}} \rightarrow h^{-1}(A^g)_{\mathcal{K}}h \quad.
\end{aligned}
\tag{6.24}
$$

Thus, $(A^g)_{\mathcal{K}}$ is a gauge–covariant derivative of the Goldstone boson field, and $(A^g)_{\mathcal{H}}$ serves as a connection in the construction of gauge–covariant derivatives of matter fields.

A kinetic term respecting local G invariance for the Goldstone boson fields is

$$
\begin{aligned}
\mathcal{L}(\phi, A) &= -\frac{1}{2} F^2 \ \mathrm{tr} \left[(A^g)_{\mathcal{K}} * (A^g)_{\mathcal{K}} \right] \\
&= -\frac{1}{2} F^2 \ \mathrm{tr} \left[(g^{-1}\partial_\mu g + g^{-1}A_\mu g)_{\mathcal{K}} (g^{-1}\partial^\mu g + g^{-1}A^\mu g)_{\mathcal{K}} \right] \quad.
\end{aligned}
\tag{6.25}
$$

When G is gauged, all of the G/H Goldstone bosons are gauge artifacts; we are free to rotate g to $\mathbf{1}$ locally by a gauge transformation. (We may, of course, contemplate gauging any proper subgroup of G, and then some physical Goldstone bosons would survive.) In the unitary gauge ($g = \mathbf{1}$), eq. (6.25) becomes a gauge field mass term; all of the "broken" gauge fields (those not in $\mathcal{H}$) acquire equal masses by eating the Goldstone bosons. Obviously, this gauge boson mass matrix is not the most general one consistent with the unbroken H gauge symmetry, and, correspondingly eq. (6.25) is not the most general gauge–invariant local operator involving two derivatives of ϕ. The Lagrangian

$$
\mathcal{L}(\phi, A) = -\frac{1}{2}(F^2)^{ab}(g^{-1}\partial_\mu g + g^{-1}A_\mu g)_{\mathcal{K}}^a (g^{-1}\partial^\mu g + g^{-1}A^\mu g)_{\mathcal{K}}^b
\tag{6.26}
$$

is locally G invariant provided only that F^2 is an H–invariant matrix. That is, there is an independent eigenvalue of F^2 corresponding to each *irreducible* representation of H that is contained in $\mathcal{K}$.

For a matter field ψ that transforms according to the representation R of H, a covariant derivative

$$D\psi \;=\; \Big\{ d \;+\; D_R\big[(A^g)_{\mathcal{H}}\big] \Big\}\psi \tag{6.27}$$

may be defined. If ψ is a left–handed Weyl fermion, then the Lagrangian

$$\begin{aligned}
\mathcal{L}(\psi, A) \;&=\; \psi^\dagger \gamma^\mu D_\mu \psi \\
&=\; \psi^\dagger \gamma^\mu \Big\{ \partial_\mu \;+\; D_R\big[(A^g_\mu)_{\mathcal{H}}\big] \Big\}\psi
\end{aligned} \tag{6.28}$$

respects local G invariance at the classical level. Gauge invariance survives quantum corrections if and only if R is an anomaly–free representation of H.

Perturbative analysis of a gauged G/H sigma model is most conveniently performed not in unitary gauge but in an 't Hooft gauge, so we will now briefly describe the implementation of the 't Hooft gauge fixing condition. Consider the expansion of eq. (6.26) in powers of the Goldstone boson field ϕ. (The expansion in powers of ϕ is adequate for carrying out chiral perturbation theory to arbitrary order in $1/F$.) It is useful to rescale ϕ by the symmetric matrix F in eq. (6.26); we write

$$g \;=\; \exp\left[\phi^a (F^{-1})^{ab} X^b \right] \quad, \tag{6.29}$$

and find, to quadratic order in the fields,

$$\begin{aligned}
\mathcal{L}(\phi, A) \;=\; &- \frac{1}{2}\partial^\mu \phi^a \partial_\mu \phi^a \;-\; F^{ab}\partial^\mu \phi_a A^b_\mu \\
&- \frac{1}{2}(F^2)^{ab} A^a_\mu A^{\mu b} \;+\; \ldots
\end{aligned} \tag{6.30}$$

Thus ϕ^a is a conventionally normalized scalar field.[*] (The ellipsis in eq. (6.30) indicates terms that are higher order in $F^{-1}\phi$.)

---

[*] Minus signs appear in eq. (6.30) because, if G is compact, A and ϕ are antihermitian fields.

The 't Hooft gauge–fixing term is now chosen to cancel the $A - \phi$ mixing in eq. (6.30); we have

$$\mathcal{L}_{\text{gaugefixing}} \;=\; -\,\frac{1}{2\alpha}\Big[(\partial^\mu A^\alpha_\mu)^2 \;+\; (\partial^\mu A^a_\mu \;+\; \alpha F^{ab}\phi^b)^2\Big] \quad , \tag{6.31}$$

corresponding to the gauge–fixing conditions

$$\partial^\mu A^\alpha_\mu T^\alpha \;+\; (\partial^\mu A^a_\mu \;+\; \alpha F^{ab}\phi^b)X^b \;=\; 0 \quad . \tag{6.32}$$

This gauge–fixing term must be accompanied by the ghost Lagrangian

$$\begin{aligned}
\mathcal{L}_{\text{ghost}} \;=\;& \partial^\mu \eta^{\dagger\alpha}[(\partial^\mu + A_\mu), \eta]^\alpha \\
&+\; \partial^\mu \eta^{\dagger a}[(\partial_\mu + A_\mu), \eta]^a \\
&+\; \alpha(F^2)^{ab}\eta^{\dagger a}\eta^b \;+\; \ldots \quad ,
\end{aligned} \tag{6.33}$$

where again we omit terms higher order in $F^{-1}\phi$. We see that, in the 't Hooft gauge, the gauge boson, Goldstone boson and ghost propagators all have gauge–artifact poles located at the eigenvalues of the matrix αF^2. These singularities cancel, of course, in calculations of gauge–invariant quantities; the cancellation is ensured by the equivalence of the 't Hooft–gauge and unitary–gauge formulations of the theory.

Having reviewed the methodology of G/H sigma models, we now return to the problem that we wish to address. Imagine that we are confronted with an anomalous theory of G gauge fields coupled to chiral fermions; the local G symmetry is linearly realized, and the fermions transform as a linear representation R that has a nontrivial anomaly. Suppose further that the representation R is anomaly–free when restricted to some nontrivial subgroup H of G. We anticipate then, in view of the discussion at the beginning of this section, that it is possible to consistently quantize this theory if all of the "broken" gauge bosons (those corresponding to generators of G that are not generators of H) are permitted to acquire mass, and even if the H gauge bosons are kept exactly massless. We expect, in fact, that we can construct a (nonrenormalizable)

theory that respects the local G symmetry by introducing Goldstone boson fields that take values in G/H. It should be possible, therefore, to cancel the anomalous variation of the fermionic effective action with a suitable local counterterm constructed from gauge fields and G/H Goldstone boson fields; we would like to see how to construct this (Wess–Zumino) term.[†]

As a first step, we consider the term

$$\Delta\Gamma_R[A,g] \;=\; \Gamma_R[A^g] \;-\; \Gamma_R[A] \quad , \tag{6.34}$$

where $\Gamma_R[A]$ is the effective action generated by integrating out the fermions in representation R. As was explained in the discussion following eq. (5.12), $\Delta\Gamma_R$ is a local functional of the gauge and Goldstone boson fields. But in contrast to the case considered in Section 5, A^g is not invariant under local G transformations realized as in eq. (6.21) and eq. (6.22); it is transformed by a local H transformation as in eq. (6.23). If R is an anomaly–free representation when restricted to H, then $\Gamma_R[A]$ is invariant under $A \to A^h$ if A takes values in $\mathcal{H}$, but not, in general, if A takes values in $\mathcal{G}$. Hence, $\Gamma_R[A^g] \;=\; \Gamma_R[A] \;+\; \Delta\Gamma_R[A,g]$ is not gauge invariant, since A^g is $\mathcal{G}$–valued.

We can, however, find a local functional that cancels the anomalous gauge variation of $\Gamma_R[A^g]$. For this purpose, we apply a result from [29]. There a local functional $B_R[A_1, A_2]$ was constructed that transforms like the difference $\Gamma_R[A_1] - \Gamma_R[A_2]$; that is,

$$\begin{aligned}B_R[A_1^\Omega, A_2^\Omega] \;-&\; B_R[A_1, A_2]\\ &= \; \Gamma_R[A_1^\Omega] \;-\; \Gamma_R[A_2^\Omega] \;-\; \Gamma_R[A_1] \;+\; \Gamma_R[A_2] \quad . \tag{6.35}\end{aligned}$$

The explicit expression for B_R is

$$B_R[A_1, A_2] \;=\; \frac{1}{48\pi^2} \int \, \mathrm{tr}\, \Big[(F_1 + F_2)(A_2 A_1 - A_1 A_2)$$

---

† The construction of Wess-Zumino terms for G/H sigma models has been discussed in [45,59,62].

$$- A_2^3 A_1 + A_1^3 A_2 + \frac{1}{2} A_2 A_1 A_2 A_1 \Big] \quad . \tag{6.36}$$

This object is a generalization of a counterterm first written down by Bardeen[43][‡]for the special case

$$G/H \;=\; [\mathrm{SU(N)}_L \;\times\; \mathrm{SU(N)}_R]/\mathrm{SU(N)}_V \quad .$$

(He used it to remove the VVV anomalies of the vector–like flavor currents in QCD.)

It follows from eq. (6.35) that the local Wess–Zumino term

$$\Gamma_R^{(\mathrm{WZ})}[A, g] \;=\; \Gamma_R[A^g] \;-\; \Gamma_R[A] \;+\; B_R[(A^g)_{\mathcal{H}}, A^g] \tag{6.37}$$

has the proper transformation property to cancel the anomalous gauge variation of $\Gamma_R[A]$. We have

$$
\begin{aligned}
\Omega \;:\; &\Gamma_R[A] \;+\; \Gamma_R^{(\mathrm{WZ})}[A, g] \;\rightarrow\; \\
&\Gamma_R[A] \;+\; \Gamma_R^{(\mathrm{WZ})}[A, g] \;+\; \Gamma_R[(A^g)_{\mathcal{H}}^h] \;-\; \Gamma_R[(A^g)_{\mathcal{H}}] \quad .
\end{aligned}
\tag{6.38}
$$

Therefore, $\Gamma_R + \Gamma_R^{(\mathrm{WZ})}$ is gauge invariant, if R is an anomaly–free representation of H.[§]

This Wess–Zumino term, unlike that constructed in Section 5 for the special case of a globally G–invariant theory, does not in general vanish in the unitary gauge where $g = \mathbf{1}$. Instead we have

$$\Gamma_R^{(\mathrm{WZ})}\Big[A, g = \mathbf{1}\Big] \;=\; B_R\Big[(A)_{\mathcal{H}}, A\Big] \quad . \tag{6.39}$$

The point is that if the local H symmetry is not spontaneously broken, then the effective action must be locally H invariant even when the Goldstone bosons are

---

‡ See also [63]

§ In [59], a different prescription was given for constructing the Wess–Zumino term. That prescription differs from eq. (6.37) by a gauge–invariant local counterterm, and in fact coincides with eq. (6.37) in the unitary gauge.

gauged away. And while $\Gamma_R[A]$ is not locally H invariant, $\Gamma_R[A] + B_R\big[(A)_{\mathcal{H}}, A\big]$ is. Indeed, the diagrams of Fig. 15 give nonzero contributions to the masses of the $\mathcal{H}$ Goldstone bosons. The diagrams of Fig. 16, in which insertions of the counterterm eq. (6.39) appear, are necessary to cancel this mass renormalization, so that the linear realization of the H gauge symmetry is preserved.

By constructing this Wess–Zumino term, and the gauge–invariant extension eq. (6.26) of the gauge boson mass matrix, we have shown that it is possible, by introducing G/H Goldstone bosons, to restore the local G symmetry in an anomalous G gauge theory whenever the fermions are in a representation that is anomaly–free with respect to H. One expects that this (nonrenormalizable) gauge–invariant version of the anomalous G gauge theory can be obtained as the low–energy limit of an underlying renormalizable anomaly–free theory in which G is spontaneously broken to H. Fermions of the underlying theory that carry a nontrivial G anomaly may acquire large masses if they are in representations that are real with respect to H. Integrating out heavy fermions and Higgs fields should result in an effective field theory of the type that we have constructed. In particular then, the Wess–Zumino counterterm eq. (6.37) should arise when the heavy fermions are integrated out.

Let us see more explicitly how the Wess–Zumino term is generated by integrating out the heavy fermions. We first note that the masses for fermions that are chiral with respect to G (but real with respect to H) must arise from Yukawa couplings to Higgs fields, and so the fermion mass matrix acquires a nontrivial dependence on the Goldstone bosons fields. If ψ' is a Weyl fermion that transforms as the linear representation R of G,

$$\Omega \; : \; \psi' \to D_R(\Omega^{-1})\psi' \quad , \tag{6.40}$$

then the locally G–invariant fermion Lagrangian has the form

$$\begin{aligned}
\mathcal{L} \; = \; &\psi'^\dagger \gamma^\mu \Big[\partial_\mu \; + \; D_R(A_\mu)\Big]\psi' \\
&- \Big[D_R(g^{-1})\psi'\Big]^T M \Big[D_R(g^{-1})\psi'\Big] \; + \; \text{h.c.} \tag{6.41}
\end{aligned}$$

Here g is the Goldstone boson field, A is the gauge field transforming as eq. (6.21–22), and M is an H–invariant matrix. Before we integrate out ψ', it is very convenient to remove the coupling of ψ' to Goldstone bosons in the mass term by performing the change of variable

$$\psi = D_R(g^{-1})\psi' \quad . \tag{6.42}$$

Of course, because of the anomaly, this change of variable has a nontrivial Jacobian J,

$$(d\psi') = J d\psi \ , \quad J = \exp\left(\Delta\Gamma_R[A,g]\right) \quad , \tag{6.43}$$

where $\Delta\Gamma_R$ is the expression in eq. (6.34).

By this change of variable, the fermionic action becomes

$$S = \int \left\{\psi^\dagger \gamma^\mu \cdot \left[\partial_\mu + D_R(A^g_\mu)\right]\psi - \psi^T M \psi + \text{h.c.}\right\} - \Delta\Gamma_R[A,g] \quad , \tag{6.44}$$

where ψ transforms as the nonlinear realization eq. (6.16) of the local G symmetry.

We may now integrate out ψ, and expand the result in powers of M^{-1}; each term in the expansion is a local functional of A^g. The effective action obtained by integrating out ψ is not gauge invariant, but its gauge variation is independent of M, because the mass term in eq. (6.44) is invariant under the nonlinearly realized G symmetry. Hence, the gauge variation of this effective action is equal to that of $\Gamma_R[A^g]$, which is the $M \to 0$ limit of the effective action. And we have seen that $\Gamma_R[A^g]$, while not gauge invariant, has the same gauge variation as the local counterterm $-B_R\left[(A^g)_{\mathcal{H}}, A^g\right]$. We therefore conclude that the result of integrating out ψ' from eq. (6.41) can be expressed as

$$\begin{aligned}
\Gamma_{\text{eff}}\left[A,g\right] = &- \Delta\Gamma_R\left[A,g\right] - B_R\left[(A^g)_{\mathcal{H}}, A^g\right] \\
&+ \Big(\text{Gauge invariant and local, when} \\
&\quad \text{expanded in powers of } M^{-1}\Big) \quad .
\end{aligned} \tag{6.45}$$

If the original renormalizable theory contains fermions in the representation R of G, plus fermions in a representation R' that cancels the G anomaly of R, then integrating out massive fermions in the representation R generates a Wess–Zumino term that cancels the anomaly of the R' fermion, as well as a gauge–invariant local functional of A and the Goldstone boson field ϕ.

(Incidentally, we may also wish to consider the more general case of an anomalous representation R of G that is anomaly–free with respect to H but is also chiral with respect to H, so that not all of the fermions in R are permitted by the surviving H symmetry to acquire masses. Then integrating out the heavy fermions generates the Wess–Zumino term as before, and the remaining massless chiral fermions remain in the low–energy effective field theory. They transform as a nonlinear realization of G, and have the locally G–invariant coupling eq. (6.28) to the Goldstone bosons and gauge fields.)

Having seen how an anomalous G gauge theory can be recast as a locally G–invariant theory, and that the ultraviolet behavior of the theory becomes transparent in an 't Hooft gauge, we are now equipped to derive an upper bound on the ultraviolet cutoff Λ in terms of the gauge boson mass matrix $(\mu^2)^{ab}$. As we saw in Section 5, this upper bound has nothing to do with anomalies. It is the mass scale at which perturbation theory breaks down in massive Yang–Mills theory; in fact, it can just as well be calculated in the *ungauged* G/H sigma model. This bound, which depends only on the eigenvalues of the matrix $(F^2)^{ab}$, is roughly of the form

$$\Lambda^2 \ \lesssim \ (4\pi F)^2 \tag{6.46}$$

where F^2 is a typical eigenvalue. Although the upper bound on Λ in a G/H sigma model is of some general interest, we will not derive its precise form here.

To conclude this section, we emphasize again that it is always possible to consistently gauge G in a G/H sigma model, as long as the fermions transform as an

anomaly–free representation of H.[⋆] It is not necessary that the fermions transform as an anomaly–free representation of G, nor indeed need they transform as a representation of G at all, if we are willing to allow G to be nonlinearly realized acting on the fermions.

The question naturally arises whether a nonrenormalizable effective field theory with nonlinearly realized local G symmetry (and linearly realized local H symmetry) may always be regarded as the low–energy limit of *some* underlying renormalizable anomaly–free theory. Such an underlying theory *can* be constructed if there is a renormalizable G gauge theory that contains the light fermions of the effective field theory as well as additional fermions that are permitted to acquire mass by the linearly realized H symmetry. And therefore the underlying renormalizable theory always exists if and only if the following proposition is true: Given any anomaly–free representation R_H of H, there is an anomaly–free representation R_G of G that decomposes with respect to H as

$$R_G \;\; \rightarrow \;\; R_H \;\; + \;\; (\text{real representation of } H) \quad . \tag{6.47}$$

It is easy to see, however, that this proposition is *false*. Consider the case $G = \mathrm{SU}(2)$ and $H = \mathrm{U}(1)$. There are, of course, anomaly–free chiral representations of $\mathrm{U}(1)$. (For example, the $Q^3 = 8$ anomaly of a charge $Q = 2$ left–handed fermion may be canceled by eight left–handed fermions with charge $Q = -1$.) But *all* representations of $\mathrm{SU}(2)$ are real with respect to $\mathrm{U}(1)$, so eq. (6.47) cannot be satisfied by any representation R_G of $\mathrm{SU}(2)$ if R_H is a nonreal representation of $\mathrm{U}(1)$. Hence, there surely do exist gauged G/H sigma models coupled to fermions that do not describe the low–energy limit of a renormalizable theory.

The question becomes somewhat more subtle if we require that the local G symmetry be linearly realized by the fermions; that is, if we require that the light fermions

---

[⋆] "Anomaly–free" meaning that both global and perturbative anomalies are absent. See Section 7.

in the effective theory transform as a representation of G, rather than just a representation of H. A gauged G/H sigma model coupled to fermions in a representation of G can always be regarded as the low–energy limit of an underlying renormalizable gauge theory if and only if the following proposition is true: Given any representation R of G that is anomaly–free with respect to $H \subset G$, there exists another representation R' of G that has the same G anomaly as R, but is *real* with respect to H. I do not know whether this proposition is true.

7. Nonperturbative Gauge Anomalies

Until now, we have considered only those gauge anomalies that can be studied in perturbation theory. But *nonperturbative* (or *global*) anomalies, which are not accessible to any order in the Feynman diagram expansion, may also threaten the consistency of a gauge theory[49]. In this section, we explain how the considerations of this paper apply to nonperturbative anomalies.

Nonperturbative anomalies potentially arise in four spacetime dimensions only if the gauge group is $G = \mathrm{Sp}(n)$, or if G contains an $\mathrm{Sp}(n)$ factor. In particular, for $G = \mathrm{SU}(2) = \mathrm{Sp}(2)$, there is an anomaly if the number of Weyl fermions in half–odd–integer spin representations of $\mathrm{SU}(2)$ is *odd*. This anomaly is associated with a topological property of $\mathrm{SU}(2)$ — namely that $\mathrm{SU}(2)$ gauge transformations on four–dimensional Euclidean spacetime (compactified to a four–sphere) are classified by a "winding number" that takes values in Z_2. A gauge transformation $\Omega(x)$ in the trivial topological class leaves the fermionic effective action $\Gamma[A]$ invariant, but if Ω is in the nontrivial class, and the fermion representation content meets the condition stated above, then Γ has the nontrivial transformation property

$$\exp\left(-\Gamma[A^\Omega]\right) \;=\; -\exp\left(-\Gamma[A]\right) \quad . \tag{7.1}$$

This nonperturbative anomaly, then, like the perturbative gauge anomaly, renders Γ ill–defined on the space of gauge field configurations modulo gauge transformations.

Consequently, we are unable to make sense of the path integral expression for the expectation value of a gauge invariant operator $\mathcal{O}$. In both numerator and denominator of

$$\langle \mathcal{O} \rangle \;=\; \frac{\int (dA)\mathcal{O}(A)e^{-S_{\text{eff}}[A]}}{\int dA e^{-S_{\text{eff}}[A]}} \quad , \tag{7.2}$$

where

$$S_{\text{eff}}[A] \;=\; S_{\text{YM}}[A] \;+\; \Gamma[A] \quad , \tag{7.3}$$

and S_{YM} is the Yang–Mills action, A and A^{Ω} give equal and opposite contributions; the path integral eq. (7.2) is 0/0 and is not well–defined.

Formally at least, we can restore the gauge invariance of Γ by introducing a Goldstone boson field $g(x)$ that takes values in SU(2), and transforms under local SU(2) transformations as

$$\Omega \;:\; g \;\to\; \Omega^{-1}g \quad . \tag{7.3}$$

Then, as in Section 5, we may add to the action the Wess–Zumino counterterm

$$\Gamma^{(\text{WZ})}[A, g] \;=\; \Gamma[A^g] \;-\; \Gamma[A] \quad . \tag{7.4}$$

It is evident that $\Gamma[A] + \Gamma^{(\text{WZ})}[A, g] = \Gamma[A^g]$ is gauge invariant[64]. Furthermore, g is a mere gauge artifact; we may set $g = \mathbf{1}$ by an appropriate choice of gauge.

In contrast to the perturbative anomaly considered in Section 5, however, the Wess–Zumino term in the case of the nonperturbative anomaly is purely topological. If we express g as $g = e^{\phi}$ and expand in powers of ϕ, eq. (7.4) vanishes to all orders in ϕ; $\Gamma^{(\text{WZ})}$ depends only on the topological class of $g(x)$, according to

$$\exp\!\Big(- \Gamma^{(\text{WZ})}[g] \Big) \;=\; 1 \quad , \quad g \text{ trivial} \quad ,$$
$$\exp\!\Big(- \Gamma^{(\text{WZ})}[g] \Big) \;=\; -1 \quad , \quad g \text{ nontrivial} \quad . \tag{7.5}$$

(It does not even depend on the gauge field A.) Of course, this dependence on g, together with the transformation property eq. (7.3), is just what is needed to cancel the anomalous transformation property of eq. (7.1).

But adding this Wess–Zumino term to the action of the theory does not, in itself, do anything to cure the disease of eq. (7.2). This is obvious, since g is merely a gauge artifact, and the Wess–Zumino term vanishes when g is gauged away. If we do introduce g and include the Wess–Zumino term in the action, then expectation values of gauge–invariant operators are given by

$$\langle \mathcal{O} \rangle \;=\; \frac{\sum_g \int (dA)\mathcal{O}e^{-S_{\text{eff}}[A,g]}}{\sum_g \int (dA)e^{-S_{\text{eff}}[A,g]}} \quad , \tag{7.6}$$

where

$$S_{\text{eff}}[A,g] \;=\; S_{\text{YM}}[A] \;+\; \Gamma[A^g] \quad , \tag{7.7}$$

and the sum runs over the two topological classes of $g(x)$. Although the effective action is now gauge invariant, eq. (7.6) is still 0/0, and hence not well–defined. In eq. (7.6), cancellations occur in both the numerator and denominator between the two contributions to the sum over g. Thus, the only difference between eq. (7.6) and eq. (7.2) is that cancellations now occur between field configurations that are not gauge equivalent. Formally restoring the gauge invariance has not really removed the anomaly.

We *can* make sense of eq. (7.2) and eq. (7.6) *if* we do not insist that the gauge fields be exactly massless. Thus, the anomalous gauge theory becomes nonrenormalizable, and the subsequent analysis is exactly as in Sections 5 and 6. In general, a theory with gauge group G that suffers from a nonperturbative anomaly can be consistently quantized, but only a subgroup H of G may be linearly realized, where H is free of both perturbative and nonperturbative anomalies. The quantum theory has an ultraviolet cutoff Λ that may not be large compared to $4\pi\mu/e$, where μ is a typical gauge boson mass.

Just as for the cases discussed in previous sections, a nonrenormalizable SU(2) gauge theory with an uncanceled global anomaly and a *massive* gauge field can arise as the low energy limit of an underlying renormalizable gauge theory that undergoes

the Higgs mechanism, when heavy fermions with a nontrivial global anomaly are integrated out.

8. Local Discrete Symmetry

The effective field theories that I have considered up to now have all contained light gauge fields, and the limits obtained on the cutoffs of these theories have been expressed in terms of gauge field masses. But it is also possible to contemplate a "gauge theory" that contains no gauge fields. In this section, I comment on how the results of this paper apply to such a theory.

Krauss and Wilczek[65] have recently called attention to continuum field theories that respect local *discrete* symmetries. A theory of this type can arise as the low–energy limit of a gauge theory with continuous gauge group G, if G is broken by the Higgs mechanism to a nontrivial discrete subgroup H. Alternatively, we can formulate a theory with discrete gauge group H directly, without appealing to an underlying theory[66–68]. In this formulation, we consider a spacetime manifold M that is not simply connected. Then quantum fields are not required to be strictly periodic around the noncontractible closed loops on M; they are permitted instead to satisfy periodicity only up to an element of H.* Thus, the discrete local H symmetry can manifest itself through a nontrivial Aharonov–Bohm effect.

For a discrete gauge theory so formulated, it is natural to ask whether cancellation of global anomalies places nontrivial constraints on the H charges carried by the fermions in the theory. But an analysis like that performed in this paper does not discern any such restriction. We note first of all that the nontrivial boundary conditions satisfied by the fields in a discrete gauge theory have no impact on the short–distance behavior of the theory, and therefore gauge invariance has no bearing on renormalizability. In contrast to a gauge theory with a continuous gauge group, there need be no spin–one fields that threaten renormalizability. In any case, the fermionic effective action is manifestly gauge invariant. The only smooth gauge transformations

---

$\star$ Of course, it makes no sense to regard the discrete H symmetry as a local symmetry unless the symmetry is *exact*.

are constants, and these have no effect on the boundary conditions satisfied by the fermions. Nor does there seem to be any reason to expect nontrivial cancellations when we sum over all possible boundary conditions. If there *are* global anomalies in theories with discrete local symmetries, then a more sophisticated analysis than ours will be needed to uncover them.

On the other hand, if global anomalies do not occur, then it seems that there exist sensible renormalizable theories with a discrete gauge group H that cannot be regarded as the low–energy limit of *any* underlying renormalizable theory with a continuous gauge group G that contains H. Consider, for example, a theory containing a single massless Weyl fermion with unit charge under $H = Z_N$. For any embedding of Z_N in U(1), this fermion has a U(1)^3 anomaly. Furthermore, this U(1)^3 anomaly cannot be canceled by any set of heavy fermions with Z_N-invariant masses. When the symmetry breakdown U(1)$\to Z_N$ occurs, a pair ψ_1, ψ_2 of Weyl fermions can acquire the mass

$$\mathcal{L}_{\text{mass}} \;=\; -M\psi_1\psi_2 + \text{h. c.} \tag{8.1}$$

only if the charges $Q_{1,2}$ of the fermions satisfy

$$Q_1 + Q_2 \;=\; \text{ integer multiple of } N \quad . \tag{8.2}$$

But then this pair of fermions has a U(1)^3 anomaly

$$Q_1^3 + Q_2^3 \;=\; (Q_1 + Q_2)(Q_1^2 - Q_1 Q_2 + Q_2^2) \tag{8.3}$$

that is also an integer mutiple of N. Since any combination of such massive fermions has a U(1)^3 anomaly proportional to N, no such combination can cancel the anomaly $(1)^3$ of the massless chiral fermion.

The existence of such renormalizable Z_N gauge theories that are not descended from any underlying renormalizable gauge theory with a continuous gauge group recalls the observation that we made at the end of Section 6 — there are consistent

(although nonrenormalizable) gauged G/H sigma models that cannot be regarded as the low–energy limit of *any* renormalizable theory with linearly realized local G symmetry.

The preceding argument assumes that the charges of all fermions in the underlying theory are integers. This assumption is necessary if we want the Z_N transformations to be single-valued acting on all of the fields of the underlying theory — that is, if we want Z_N to be a subgroup of U(1). We could relax this assumption and demand only that the Z_N transformations are well-defined when acting on the light degrees of freedom that survive after heavy fermions are integrated out. Then the heavy fermions can have non-integer charges, and our conclusion, that the Z_N gauge theory cannot be obtained as the low–energy limit of an underlying U(1) gauge theory, might be evaded.

9. Conclusions

We have shown that gauge theories in four spacetime dimensions with anomalous fermion content can be consistently quantized. Unitarity and covariance can be maintained if we are willing to allow at least some of the gauge fields to have nonvanishing masses. The resulting quantum theory is nonrenormalizable, and so must be equipped with a finite ultraviolet cutoff Λ. For a given anomalous theory, there is an upper bound on Λ beyond which the locality of the theory breaks down. This upper bound can be expressed in terms of gauge boson masses.

The upper limit on Λ can be estimated by a perturbative analysis of the anomalous gauge theory. We may also interpret the bound on Λ in terms of an underlying renormalizable field theory, from which the anomalous gauge theory may be obtained by integrating out heavy fermion and Higgs fields. We have emphasized that, with regard to the upper bound on Λ, a qualitative distinction may be made between anomalous Abelian and non–Abelian gauge theories. Because a massive anomaly–free Abelian gauge theory is renormalizable, the bound on Λ in an Abelian gauge theory depends both on the gauge field mass and on the strength of the anomaly. But because

massive Yang–Mills theory is nonrenormalizable even in the absence of anomalies, the bound on Λ in a non–Abelian gauge theory depends only on gauge field masses, at least if the gauge coupling is sufficiently weak. From the viewpoint of an underlying renormalizable theory, the Abelian and non–Abelian cases are distinguished by the property that electric charge is unquantized in an Abelian gauge theory.

For the case of an anomalous U(1) guage theory, the upper bound on Λ was estimated in Section 2 to be

$$\Lambda \lesssim \mu/\left(\frac{e^3}{64\pi^3}\,|\text{tr}\,(Q^3)|\right) \quad, \tag{9.1}$$

where μ is the mass of the U(1) gauge boson, e is the corresponding gauge coupling, and Q is the U(1) generator acting on the left-handed Weyl fermions of the theory. Other limits on Λ, associated with mixed and gravitational anomalies, were obtained in eq. (4.13) and eq. (4.18).

The proper way to derive an upper bound on Λ would be to study the exact renormalization group flow in the effective field theory. That being beyond my ability, I have resorted to perturbative approximations. Thus, my quantitative conclusions should not be taken too seriously. Even the qualitative conclusions are not above suspicion. It is certainly conceivable that the ultraviolet behavior of the sigma models studied here is actually governed by a nontrivial fixed point of the renormalization group, even though there is no hint of such a fixed point in perturbation theory. If so, it might be possible in principle to push the cutoff Λ of the effective field theory as high as one pleases, without encountering any breakdown of quasilocality or unitarity. Another shortcoming of my perturbative approach is that by focusing on the propagating degrees of freedom of the effective field theory, I have ignored all considerations of global topoogy. The role of possible global gauge excitations [69] in anomalous gauge theories deserves further attention.[*]

I wish to conclude with some remarks about the "phenomenological" implications of these results. The analysis described here addresses the following question: Suppose

---

[*] For a recent discussion, see [70].

that low–energy phenomenology were found to be well–described by an effective field theory that appeared to have gauge anomalies. At what energy scale, then, must new physics necessarily be encountered?

In the case of the standard model, the anomalies of the known light fermions (including the top quark) are found to cancel. In fact, the requirement of anomaly cancellation is sometimes invoked as an explanation for the peculiar U(1) hypercharge assignments of the quarks and leptons[71,72]. My arguments show that this requirement can actually be evaded; a tiny amount of uncanceled anomaly can be tolerated if accompanied by new physics at a correspondingly high scale. Still, the anomaly–cancellation argument that explains the U(1) hypercharge assignments must be regarded as very well founded, because the mass of the photon is known to be extraordinarily small. The constraint most easily evaded is

$$\text{tr } Q_{\text{Y}} \ = \ 0 \quad , \tag{9.2}$$

where Q_{Y} is weak hypercharge, which is the condition for the cancellation of the gravitational anomaly. From the (conservative) observed bound[73] on the photon mass μ,

$$\mu \ \lesssim \ (10^6 \text{ km})^{-1} \ \sim \ 10^{-25} \text{ GeV} \quad , \tag{9.3}$$

together with eq. (4.18), we infer the upper bound on the scale of new physics

$$\Lambda \ \lesssim \ 10^5 \text{ GeV}/(|\text{tr } (Q_{\text{Y}})|)^{1/3} \quad . \tag{9.4}$$

Thus, a gross failure of anomaly cancellation could in principle be accommodated if the standard model were embedded in an underlying theory that undergoes the Higgs phenomenon at a mass scale less than or of order 10^5 GeV.[†] Of course, for the Higgs mechanism at the scale 10^5 GeV to generate a photon mass of order 10^{-25} GeV would require an exceedingly tiny charge for the Higgs field, but such a small charge is at

---

† This was observed by Alvarez–Gaumé and Witten[41].

least a theoretical possibility. (This possibility is excluded if $U(1)_Y$ is embedded in a unified theory with a semisimple gauge group, but in that event, eq. (9.2) would be an automatic consequence of unification, and there would be no need to invoke anomaly cancellation in order to explain it.)

When we consider physics beyond the standard model, we may contemplate the possibility that additional Z^0's will someday be discovered at higher energies, and it will be necessary to extend the gauge group of the standard model by introducing new $U(1)$ gauge symmetries. Perhaps we will find that the known light fermions carry $U(1)$ charges that are apparently anomalous. In this situation, it will be natural to speculate about the masses of the as yet unknown fermions that cancel these anomalies. Then eq. (9.1), eq. (4.13), and eq. (4.18) may be cited as upper bounds on the mass scale Λ of this new physics.

ACKNOWLEDGEMENTS

REFERENCES

1. D. J. Gross and R. Jackiw, Phys. Rev. **D6** (1972) 477; C. Bouchiat, J. Illiopoulos, and Ph. Meyer, Phys. Lett. **38B** (1972) 519.

2. G. 't Hooft, Nucl. Phys. **B33** (1971) 173; **B35** (1971) 167; G. 't Hooft and M. Veltman, Nucl. Phys. **B44** (1972) 189; **B50** (1972) 318.

3. K. Wilson and J. Kogut, Phys. Rep. **12C** (1974) 75.

4. J. Polchinski, Nucl. Phys. **B231** (1984) 269.

5. B. J. Warr, Ann. Phys. **183** (1988) 1, 59.

6. L. Faddeev, in *Supersymmetry and its applications*, ed. G. W. Gibbons, S. W. Hawking, and P. K. Townsend (Cambridge University Press, Cambridge, 1986); L. D. Faddeev and S. L Shatashvili, Phys. Lett. **167B** (1986) 225.

7. N. V. Krasnikov, JETP Lett. **41** (1985) 586.

8. J. Wess and B. Zumino, Phys. Lett. **37B** (1971) 95.

9. E. Witten, Nucl. Phys. **b223** (1983) 422.

10. C. Becchi, A. Rouet, and R. Stora, Comm. Math. Phys. **42** (1975) 127; Ann. Phys. **98** (1976) 287; I. V. Tyutin, Lebedev Report FIAN n. 39 (1975) unpublished.

11. R. Jackiw and R. Rajaraman, Phys. Rev. Lett. **54** (1985) 1219, 2060(E); R. Rajaraman, Phys. Lett. **154B** (1985) 305.

12. O. Babelon, F. A. Shaposnik, and C. M. Viallet, Phys. Lett. **B177** (1986) 385.

13. K. Harada and I. Tsutsui, Phys. Lett. B183 (1987) 311.

14. S. G. Rajeev, "Quantization of an anomalous gauge theory," MIT Report CTP 1405 (1986), unpublished.

15. A. J. Niemi and G. W. Semenoff, Phys. Rev. Lett.**56** (1986) 1019.

16. R. Rajaraman, Phys. Lett. **B184** (1987) 369.

17. G. Siopsis, "Building new chiral gauge theories," Caltech preprint CALT-68-1335 (1986) unpublished.

18. G. Thompson and R. Zhang, Mod. Phys. Lett. **A2** (1987) 869.

19. F. Kaschluhn, Annalen Phys. **45** (1988) 539.

20. S. Miyake and K. Shizuya, Mod. Phys. Lett. **A4** (1989) 2675.

21. A. Andrianov, A. Bassetto, and R. Soldati, Phys. Rev. Lett. **63** (1989) 1554.

22. T. D. Kieu, Phys. Lett. **B242** (1990) 234.

23. S. Aoki, "Perturbative analysis of anomalous chiral QED," Stony Brook preprint ITP-SB-90-3 (1990).

24. M. B. Einhorn and G. J. Goldberg, Phys. Rev. Lett. **57** (1986) 2115.

25. R. D. Ball, Phys. Lett. **B183** (1987) 315.

26. T. Sterling and M. Veltman, Nucl. Phys. **B189** (1981) 557.

27. E. D'Hoker and E. Farhi, Nucl. Phys. **B248** (1984) 59, 77.

28. J. Preskill and M. B. Wise, unpublished (1987).

29. L. Alvarez–Gaumé and P. Ginsparg, Ann. Phys. **161** (1985) 423; **171** (1986) 233(E).

30. S. L. Adler, Phys. Rev. **177** (1969) 2426; J. S. Bell and R. Jackiw, Nuovo. Cim. **60A** (1969) 47.

31. W. A. Bardeen and B. Zumino, Nucl. Phys. **B244** (1984) 421.

32. K. Fujikawa, Phys. Rev. Lett. **42** (1979) 1195.

33. L. D. Faddeev and V. N. Popov, Phys. Lett. **25B** (1967) 30.

34. M. B. Green and J. H. Schwarz, Phys. Lett. **149B** (1984) 117.

35. E. Witten, Phys. Lett. **149B** (1984) 351; M. Dine, N. Seiberg, and E. Witten, Nucl. Phys. **B289** (1987) 589.

36. J. A. Harvey and S. G. Naculich, Phys. Lett. **B217** (1989) 231.

37. A. Cohen, H. Georgi, and E. H. Simmons, Phys. Rev. **D38** (1988) 405.

38. K. G. Wilson, Phys. Rev. **D3** (1971) 1818.

39. L. Susskind, Phys. Rev. **D20** (1979) 2619.

40. B. Zumino, in *Relativity, groups, and topology II*, ed. B. S. DeWitt and R. Stora (North-Holland, Amsterdam, 1984); B. Zumino, Y.–S. Wu, and A. Zee, Nucl. Phys. **B239** (1984) 477.

41. L. Alvarez–Gaumé and E. Witten, Nucl. Phys. **B234** (1983) 269.

42. D. Delbourgo and A. Salam, Phys. Lett. **40B** (1972) 381; T. Eguchi and P. Freund, Phys. Rev. Lett. **37** (1976) 1251.

43. W. A. Bardeen, Phys. Rev. **184** (1969) 1848.

44. A. Manohar and G. Moore, Nucl. Phys. **B243** (1984) 55.

45. H. Sonoda, Caltech Ph. D. thesis (1985), unpublished.

46. P. Galison, Nucl. Phys. **B232** (1984) 26.

47. M. Aizenman, Phys. Rev. Lett. **47** (1981) 1, 886(E); Comm. Math. Phys. **86** (1982) 1; J. Fröhlich, Nucl. Phys. **B200** (1982) 281.

48. R. Dashen and H. Neuberger, Phys. Rev. Lett. **50** (1983) 1897.

49. E. Witten, Phys. Lett. **117B** (1982) 324.

50. P. Sikivie, L. Susskind, M. Voloshin, and V. Zakharov, Nucl. Phys. **B173** 91980) 189.

51. A. A. Slavnov, Nucl. Phys. **B31** (1971) 301.

52. T. D. Lee and G. C. Wick, Nucl. Phys. **B9** (1969) 209; Phys. Rev. **D2** (1970) 1033.

53. D. G. Boulware and D. J. Gross, Nucl. Phys. **B233** (1984) 1.

54. P. Hasenfratz, Nucl. Phys. **B321** (1989) 139.

55. S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177** (1969) 2239; C. Callan, S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177** (1969) 2247.

56. G. Moore and P. Nelson, Phys. Rev. Lett. **53** (1984) 1519; Comm. Math. Phys. **100** (1985) 83.

57. A. Manohar, G. Moore, and P. Nelson, Phys. Lett. **152B** (1985) 68.

58. J. Bagger, D. Nemeschansky, and S. Yankielowicz, Nucl. Phys. **B262** (1985) 478.

59. L. Alvarez–Gaumé and P. Ginsparg, Nucl. Phys. **B262** (1985) 439.

60. I. S. Gerstein, R. Jackiw, B. W. Lee, and S. Weinberg, Phys. Rev. **D3** (1971) 2486.

61. A. C. Redfield, Phys. Lett. **109B** (1982) 311.

62. Y.–S. Wu, Phys. Lett. **153B** (1985) 70.

63. J. Manes, Nucl. Phys. **B250** (1985) 369.

64. R. Percacci, Mod. Phys. Lett. **A2** (1987) 977.

65. L. M. Krauss and F. Wilczek, Phys. Rev. Lett. **62** (1989) 1221.

66. L. M. Krauss, Gen. Rel. Grav. **22** (1990) 253.

67. T. Banks, Nucl. Phys. **B323** (1989) 90.

68. J. Preskill and L. M. Krauss, Nucl. Phys. **B341** (1990) 50.

69. E. Witten, Comm. Math. Phys. **117** (1988) 353.

70. Y.–S. Wu and W.–D. Zhao, Phys. Rev. Lett. **66** (1990) 675.

71. M. Wise, in *Physics in Collision*, ed. A. Seiden (Éditions Frontières, Gif-sur-Yvette, 1984).

72. J. A. Minahan, P. Ramond, and R. C. Warner, Phys. Rev. **D41** (1990) 715.

73. L. Davis, A. S. Goldhaber, and M. M. Nieto, Phys. Rev. Lett. **35** (1975) 1402.

FIGURES

1. Diagrams that produce the anomalous part of the effective action $\Gamma[A]$. External lines are gauge fields and internal lines are chiral fermions.

2. Diagram that generates a kinetic term for the θ field. Boxes are insertions of the Wess-Zumino term eq. (2.6).

3. Diagrams that, when combined with the diagram in Fig. 2, form a gauge-invariant set.

4. Diagrams that generate the operator eq. (2.23).

5. A diagram that contributes to the renormalization of the scalar field b.

6. A diagram that generates a six-point coupling for b.

7. A diagram that renormalizes the four-point coupling for b.

8. A diagram that contributes to the mass of the photon. The internal line represents a fermion that has acquired mass from the Yukawa coupling eq. (3.7).

9. Diagrams that renormalize the gauge-invariant kinetic term of the Higgs field ϕ. The indices 1,2 label the Weyl fermions that appear in eq. (3.7).

10. A diagram that contributes to the mass of the SU(2) guage boson, in a theory with a nontrivial mixed anomaly. Here 1 denotes the U(1) gauge field, and 2 denotes the SU(2) gauge field.

11. Diagrams that cancel the gauge field mass generated by the diagram in Fig. 10. The box is an insertion of the counterterm eq. (4.8).

12. Diagrams that generate the gravitational anomaly. The wavy line is a U(1) gauge field, and the looped lines are gravitons.

13. Gauge-invariant set of diagrams that generate a photon mass, in a theory with a nontrivial gravitational anomaly.

14. Diagram that renormalizes the ϕ field in the sigma model.

15. Diagrams that contribute to gauge field masses in the gauged G/H sigma model. Here $\mathcal{H}$ denotes H gauge fields, and $\mathcal{K}$ denotes gauge fields that couple to the broken generators of G.

16. Diagrams that cancel the contribution due to the diagrams in Fig. 15. The box is an insertion of the counterterm eq. (6.39).