4.3 Chiral (or Axial Vector) Anomalies

The fact that classical symmetries need not survive quantization is now a well-established, but still poorly understood fact. The whole subject of symmetry plays a much more vigorous role in a quantum theory than it ever did in classical dynamics. For aesthetic and practical reasons, we prefer theories with a high degree of symmetry, but because Nature is asymmetric, we must also account for this symmetry breaking. The oldest and most primitive idea for symmetry breaking is that of "approximate" symmetries. One supposes that there are terms in the Lagrangian that violate the symmetry, but they are "small". More refined is the concept of spontaneous symmetry breaking, introduced by W. Heisenberg in condensed-matter physics, and extended by him as well as by J. Goldstone and Y. Nambu to the particle-physics domain. Here the dynamical equations are completely symmetric, but energetic considerations of stability indicate that the ground state is asymmetric. As is well known, these ideas are realized in the modern theory of low-energy processes involving pseudo-scalar mesons, principally the pion and in the unified models for weak, electromagnetic as well as (speculatively) strong interactions. Anomalous breaking of symmetries — the third, most subtle mechanism — arises from quantum mechanical effects, in a way whose fundamental origin remains obscure. Certainly there are no energetic or stability considerations as in spontaneous breaking. Our only clue comes from perturbation theory: there does not exist a regularization procedure which respects the anomalously broken symmetries. In addition to the scale and conformal symmetries, on whose anomalous breaking I have already commented, it is the chiral fermion symmetries that are anomalously broken and that possess a rich topological structure, even in flat space. Both symmetries are dimension-specific hence dimensional regularization breaks them. Both symmetries rely on zero-mass fields, hence Pauli-Villars regularization breaks them as well.

Nevertheless, there is good reason to believe that anomalies are not obscure consequences of problems with perturbation theory, but reflect a deep fact about Nature, which when understood will surely illuminate a whole complex of related ideas: chirality, spontaneous mass generation, spontaneous symmetry breaking and the reasons for parity violation. Moreover, as I shall show in a two-dimensional example, the occurrence of anomalies can be established from general principles, with no recourse to perturbation theory. In that example, the anomaly will also be responsible for spontaneous mass generation. In higher dimensions, we do not have such explicit construction of the anomalies, but we shall present non-perturbative/topological arguments for the existence of some of them.

Thus we expect that anomalies are a true aspect of quantum mechanical Nature, and their prevalence in all branches of physics gives support for this: consequences of scale and chiral anomalies are widely proliferated in particle physics [61]; scale (= trace) anomalies are widely studied in gravity theory [13], even though chiral anomalies have not yet had an impact there [62]; finally in condensed-matter physics, scale anomalies lead to the understanding of critical phenomena [63] and chiral anomalies are just beginning to enter the field [64].

Although the subject is contemporary [65], the idea that symmetries may be broken by quantum effects possesses a prehistory. Before the neutrino hypothesis, some speculated that the β-decay spectrum indicates a quantum mechanical violation of energy conservation. Also before gauge invariant Pauli-Villars regularization was developed for quantum electrodynamics, there was some question whether electromagnetic gauge invariance could be maintained in a quantum field theory. Both puzzles were ultimately resolved, and symmetries were maintained, but the idea that quantum effects can eliminate a classical conservation law has survived and is realized in the anomaly phenomenon.

Since anomalies arise from the unavoidable infinities of relativistic and local field theory, specifically when fermions are involved (Dirac's negative energy sea is one example) the resulting formulas for the anomalies reflect the ambiguities which arise when infinities are regulated. Consequently, there is a certain amount of arbitrariness in the expressions. In perturbation theory, the source of ambiguity comes from the fact that the renormalization rules for perturbation theory allow adjusting the value of any graph by arbitrary local functions of the coordinates (polynomials in the momenta). These local terms can also modify current divergences.

The arbitrariness is somewhat limited when it is realized that three different types of currents are under discussion:

(1) The most important current in a gauge theory is the source current J_a^μ to which gauge fields couple. If it is possible, the regularization and renormalization scheme must be chosen in such a way that this current be covariantly conserved. Moreover, all other physical quantities must be gauge invariant, and the regularization procedure must define them in a gauge-invariant fashion. If it is impossible to maintain source current conservation, the theory loses gauge invariance, and it is therefore rejected [66].

(2) There may also be in the model Noether symmetry currents j^μ which are classically conserved, but are not coupled to gauge fields. Their regularized version should be defined consistently with the gauge principle, but if they

fail to be conserved, the theory need not be rejected, although its symmetry will be reduced.

(3) Finally there may be "partially" conserved currents, whose formal divergences are "small", reflecting an "approximate" symmetry. The anomalies associated with these currents are the most arbitrary, since it may not be possible to separate unambiguously a quantum addition to a nonzero divergence which is already present classically. Nevertheless the regularization should be gauge invariant.

Here I shall discuss mainly chiral anomalies of the first two categories. Before examining them in realistic four-dimensional models, let us look at some two-dimensional Abelian models where much the same phenomena can be seen in a simpler setting where results can be established without perturbation theory. (There are no anomalous divergences of chiral currents in three- or any other odd-dimensional space-time; however, there are other chiral anomalies in odd-dimensional field theories; see section 5.2.)

In two dimensions, Dirac fields are two-component objects (aside from any further degrees of freedom associated with internal symmetry) and Dirac matrices may be chosen to be Pauli matrices:

$$\gamma^0 = \sigma^1 \quad , \qquad \gamma^1 = \mathrm{i}\sigma^2 \quad , \qquad \gamma_5 = -\mathrm{i}\sigma^3 \qquad . \tag{4.7}$$

A peculiar property of these two-dimensional matrices, which leads to all our results, is that the axial vector is dual to the vector.

$$\epsilon^{\mu\nu}\gamma_\nu = \mathrm{i}\gamma^\mu\gamma_5 \qquad , \qquad \epsilon^{01} = 1 = -\epsilon_{01} \qquad . \tag{4.8}$$

Hence the axial vector current $j_5^\mu \equiv \hbar\bar{\psi}\mathrm{i}\gamma^\mu\gamma_5\,\psi$ is dual to the vector current $j^\mu \equiv \hbar\bar{\psi}\gamma^\mu\psi$,

$$j_5^\mu = \epsilon^{\mu\nu} j_\nu \quad , \tag{4.9}$$

But now it follows that in a dynamically non-trivial theory, both currents cannot be conserved. To see this, consider the vacuum correlation function of two vector currents, whose most general, Poincaré-invariant, form is

$$\langle j^\mu(x) j^\nu(y)\rangle = g^{\mu\nu}\Pi_1(x-y) - \frac{\partial^\mu\partial^\nu}{\Box}\Pi_2(x-y)$$
$$+\left(\frac{\partial^\mu\epsilon^{\nu\alpha}\partial_\alpha}{\Box} + \frac{\partial^\nu\epsilon^{\mu\alpha}\partial_\alpha}{\Box}\right)\Pi_3(x-y) \quad . \tag{4.10a}$$

The axial vector-vector correlation function is determined by the above

$$\langle j_5^\mu(x) j^\nu(y)\rangle = \epsilon^{\mu\alpha}\langle j_\alpha(x) j^\nu(y)\rangle$$

$$= \epsilon^{\mu\nu}\,\Pi_1(x-y) - \epsilon^{\mu\alpha}\,\frac{\partial_\alpha\partial^\nu}{\Box}\,\Pi_2(x-y)$$

$$- \left(g^{\mu\nu} - 2\,\frac{\partial^\mu\partial^\nu}{\Box}\right)\Pi_3(x-y) \quad . \qquad (4.10b)$$

Vector current conservation requires $\Pi_1 - \Pi_2 = 0$, and $\Pi_3 = 0$, but axial vector conservation would be obtained only if $\Pi_2 = 0$ and $\Pi_3 = 0$. The two are incompatible for a non-trivial theory.

Detailed calculation reveals that the one-loop graph contributing to the two-current correlation function cannot have its logarithmic divergence regulated so that both vector and axial vector vertices are conserved [67]. Moreover, when these currents arise in an Abelian gauge theory with massless fermions — if the coupling is vector-like one is speaking of the well-known Schwinger model of two-dimensional massless electrodynamics [68] — the functional determinant may be explicitly computed. Three couplings to Dirac fermions may be considered: vector, pseudovector, and chiral, giving rise to the following determinants:

$$\Delta_1(A) = \det(\not\partial + \mathrm{i}e\not A) \quad ,$$

$$\Delta_2(A) = \det(\not\partial - e\gamma_5\not A) \quad ,$$

$$\Delta_3(A) = \det(\not\partial + \mathrm{i}e\,\tfrac{1}{2}\not A(1 \pm \mathrm{i}\gamma_5)) \quad . \qquad (4.11)$$

By virtue of (4.8), the last two determinants may be rewritten in the same form as the first, since $\mathrm{i}\gamma^\mu\gamma_5 A_\mu = \gamma^\mu\{-\epsilon_{\mu\nu}A^\nu\}$ and $\frac{1}{2}\gamma^\mu(1 \pm \mathrm{i}\gamma_5)A_\mu = \gamma^\mu\{\frac{1}{2}(g_{\mu\nu} \mp \epsilon_{\mu\nu})A^\nu\}$. Note that $\frac{1}{2}(g^{\mu\nu} \mp \epsilon^{\mu\nu})$ is a projection operator; only one component of A^μ couples. Since the first determinant is known [68], closed forms can be given for all three:

$$-\mathrm{i}\hbar\ln\Delta_1(A) = \frac{\hbar e^2}{2\pi}\int \mathrm{d}x\mathrm{d}y$$

$$\times\; A^\mu(x)\left[g_{\mu\nu}a_1 - \frac{\partial_\mu\partial_\nu}{\Box}\right]\delta(x-y)A^\nu(y) \quad , \qquad (4.12)$$

$$-i\hbar \ln \Delta_2(A) = \frac{\hbar e^2}{2\pi}\int dx\, dy$$

$$\times\ A^\mu(x)\left[g_{\mu\nu}a_2 - \frac{\partial_\mu \partial_\nu}{\square}\right]\delta(x-y)A^\nu(y)\ , \tag{4.13}$$

$$-i\hbar \ln \Delta_3(A) = \frac{\hbar e^2}{8\pi}\int dx\, dy A^\mu(x)\left[g_{\mu\nu} - 2\frac{\partial_\mu \partial_\nu}{\square}\right.$$

$$\left. \mp \frac{\partial_\mu \epsilon_{\nu\alpha}\partial^\alpha}{\square} \mp \frac{\partial_\nu \epsilon_{\mu\alpha}\partial^\alpha}{\square}\right]\delta(x-y)A^\nu(y)\ . \tag{4.14}$$

Here a_1 and a_2 are coefficients of local terms which are undetermined by the one-loop graph — they may be fixed at will. For the vector case, gauge invariance dictates that a_1 be set to unity. The determinant is then gauge invariant; the vector current is the source current and it is conserved. But the axial vector current — a Noether current for the axial vector symmetry of massless fermions and obtained by operating with $\epsilon^{\mu\nu}\delta/\delta A_\nu$ on the determinant — possesses an anomaly [67].

$$-i\hbar \ln \Delta_1(A) = -\frac{\hbar e^2}{4\pi}\int dx F^{\mu\nu}\frac{1}{\square}F_{\mu\nu} = \frac{\hbar e^2}{2\pi}\int dx\, {}^*F\frac{1}{\square}{}^*F\ ,$$

$$\partial_\mu J^\mu = 0, \qquad \partial_\mu j_5^\mu = -\frac{\hbar e}{2\pi}\epsilon^{\mu\nu}F_{\mu\nu} = -\frac{\hbar e}{\pi}{}^*F\ . \tag{4.15}$$

(Here and below, the factor $\hbar$ in the anomalous divergence is present because the current is defined with that factor as well. Also $\hbar$ reflects the "one-loop" nature of the anomaly.) A similar gauge-invariant expression for the determinant emerges in the second case with a_2 set to unity. A conserved axial vector current exists, but now the anomaly is in the vector current — the Noether fermion number current:

$$-i\hbar \ln \Delta_2(A) = -\frac{\hbar e^2}{4\pi}\int dx F^{\mu\nu}\frac{1}{\square}F_{\mu\nu} = \frac{\hbar e^2}{2\pi}\int dx\, {}^*F\frac{1}{\square}{}^*F\ ,$$

$$\partial_\mu J_5^\mu = 0\ , \qquad \partial_\mu j^\mu = \frac{\hbar e}{2\pi}\epsilon^{\mu\nu}F_{\mu\nu} = \frac{\hbar e}{\pi}{}^*F\ . \tag{4.16}$$

In both cases, the anomalously non-conserved current may be defined gauge invariantly, and the anomaly is gauge invariant, given by twice the Pontryagin density. Finally, in the third case, the local term cannot be adjusted at will because the kernel must have the projection properties of $(g_{\mu\nu} \mp \epsilon_{\mu\nu})$ so that only one component of A_μ be present in (4.14). This is assured by the expression exhibited, which however is not gauge invariant. Rather one finds

$$-i\hbar \ln \Delta_3(A) = \frac{\hbar e^2}{4\pi} \int d^2x \left({}^*F \frac{1}{\Box} {}^*F \mp {}^*F \frac{1}{\Box} \partial_\mu A^\mu - \frac{1}{2} A^2 \right),$$

$$\partial_\mu J^\mu = \mp \frac{\hbar e}{4\pi} ({}^*F \mp \partial_\mu A^\mu) = \frac{\hbar e}{4\pi} (g^{\mu\nu} \mp \epsilon^{\mu\nu}) \partial_\mu A_\nu \quad . \tag{4.17}$$

The anomaly is not even gauge invariant, though it can be written as a total divergence. Since here the source current is not conserved, the quantized gauge theory has lost gauge invariance [66, 69].

Exercise 4.1. Verify from the explicit formulas for the appropriate determinants that the anomalous divergence equations are given by (4.15), (4.16) and (4.17).

Let us observe that in the first two (consistent) examples the anomaly is expressed by the two-dimensional Pontryagin density (3.51). Moreover, one may also show that the massless vector meson spontaneously acquires a mass without the intervention of scalar symmetry breaking fields. A straightforward argument, applied to the vector Schwinger model (first example above), makes use of the topological quantity that emerges in connection with the axial vector anomaly. Consider the gauge field equation

$$\partial_\mu F^{\mu\nu} = e J^\nu \quad . \tag{4.18a}$$

By contracting with $\epsilon_{\nu\alpha}$ this becomes

$$\partial^\mu {}^*F = e j_5^\mu \quad . \tag{4.18b}$$

Taking a second divergence and using the anomalous conservation equation for j_5^μ yields a free equation for ${}^*F$ which shows explicitly that the gauge field is massive [70]. (Recall that the two-dimensional coupling constant has the dimension of mass, in units where $\hbar$ and the velocity of light are dimensionless.)

$$\Box\, {}^*F = e\partial_\mu j^\mu_5 = -\frac{\hbar e^2}{\pi}\, {}^*F \quad . \tag{4.18c}$$

While we shall see another topological mechanism for vector meson mass generation in three space-time dimensions, no similarly elegant result has yet been established in four dimensions.

If in the models with chiral couplings we drop the components of the Dirac field that do not participate in the interaction, we can write the respective fermionic Lagrangians $\mathscr{L}_+ (\mathscr{L}_-)$ solely in terms of right (left) spinors coupled to $\frac{1}{2}(g_{\mu\nu} - \epsilon_{\mu\nu})A^\nu$ whose only non-vanishing component is $A^+ = (A^0 + A^1)/\sqrt{2}$ [respectively, $\frac{1}{2}(g_{\mu\nu} + \epsilon_{\mu\nu})A^\nu$ with non-vanishing component $A^- = (A^0 - A^1)/\sqrt{2}$].

$$\mathscr{L}_\pm = \hbar\bar{\psi}_\pm(i\partial\!\!\!/ - eA\!\!\!/)\psi_\pm \quad . \tag{4.19}$$

The fermion determinants, obtained from (4.14), are respectively

$$-i\hbar \ln \Delta_3^+(A) = -\frac{\hbar e^2}{4\pi}\int A^+ \frac{\partial_+}{\partial_-} A^+ \quad , \tag{4.20a}$$

$$-i\hbar \ln \Delta_3^-(A) = -\frac{\hbar e^2}{4\pi}\int A^- \frac{\partial_-}{\partial_+} A^- \quad . \tag{4.20b}$$

On the other hand, the vector theory (Schwinger model) is described by a fermionic Lagrangian which is the sum of left and right Lagrangians.

$$\mathscr{L}_V = \mathscr{L}_+ + \mathscr{L}_- \quad . \tag{4.21}$$

In this model, the conservation of the vector current, as dictated by gauge invariance, and the anomalous non-conservation of the axial vector current, see (4.15), indicate that separately the right and left currents are not conserved, but their sum is. Why this should be the case, in spite of the fact that the formal Lagrangian in (4.21) exhibits no interaction between left and right spinors, is the surprise of the chiral anomaly. The puzzle is resolved when it is appreciated that the gauge invariant effective Lagrangian given by the fermion determinant (4.12) with $a_1 = 1$, is not merely the sum of left and right terms, neither of which is separately gauge invariant nor is their sum. Rather, to insure gauge invariance a contact term which couples left to right must be added.

$$-i\hbar \ln \Delta_1(A) = -i\hbar \ln \Delta_3^+(A)\Delta_3^-(A) + \frac{\hbar e^2}{2\pi^2}\int A^+ A^- \quad . \qquad (4.22)$$

This shows how the anomaly in the axial Noether current is forced by demanding vector gauge invariance. Also, we see that, contrary to assertions in the literature [62], the determinant for Dirac fermions is not merely the product of determinants for Weyl fermions of each chirality.

Finally, we remark that very similar results hold in two-dimensional non-Abelian fermionic models, since also for these the fermionic determinant may be evaluated, not explicitly in terms of the vector potentials as in (4.12) and (4.14), but rather in terms of non-local matrix functionals of the vector potentials defined by $A_\pm = g_\pm^{-1}\,\partial_\pm\, g_\pm$, where $g_\pm$ are group elements [71].

Turning now to anomalies in four dimensions, the following summarizes the results of the last decade's research [72]. While it is impossible to evaluate the functional determinant exactly, the anomalous graphs have been identified — they are the reflection non-symmetric triangle graphs, involving one or three axial vertices [73]. No other graphs introduce new structures, except that one may need to adjust them so that the anomalous divergence possesses a preferred form [74]. Consequently purely vectorial gauge theories, like quantum chromodynamics are gauge invariant. However, when the fermions are massless, the gauge invariant, group singlet axial vector Noether current j_5^μ is not conserved. Rather it satisfies

$$\begin{aligned}\partial_\mu j_5^\mu &= \frac{\hbar}{8\pi^2}\,\mathrm{tr}\,{}^*\mathscr{F}^{\mu\nu}\mathscr{F}_{\mu\nu} \\ &= \frac{\hbar}{8\pi^2}\,\partial_\mu \epsilon^{\mu\alpha\beta\gamma}\,\mathrm{tr}\,(\mathscr{F}_{\alpha\beta}\mathscr{A}_\gamma - \tfrac{2}{3}\mathscr{A}_\alpha\mathscr{A}_\beta\mathscr{A}_\gamma) \\ &= \frac{\hbar}{4\pi^2}\,\partial_\mu \epsilon^{\mu\alpha\beta\gamma}\,\mathrm{tr}\,(\mathscr{A}_\alpha\partial_\beta\mathscr{A}_\gamma + \tfrac{2}{3}\mathscr{A}_\alpha\mathscr{A}_\beta\mathscr{A}_\gamma) \quad . \qquad (4.23)\end{aligned}$$

For one fermion in the fundamental representation, this is twice the Pontryagin density apart from $-\hbar$. Since the anomaly is a total divergence, [see (3.44)] a conserved axial vector current can be defined, but it is not gauge invariant:

$$\tilde{j}_5^\mu = j_5^\mu + 2\hbar\mathscr{C}^\mu \Longleftrightarrow \partial_\mu \tilde{j}_5^\mu = 0 \quad . \qquad (4.24a)$$

The charge constructed from the conserved current is time independent and invariant against small gauge transformations. Under a large gauge transformation, it changes by $\hbar$ times twice the winding number of the transformation, since the anomalous addition to the charge is twice $\hbar W(A)$, see (3.45), whose gauge transformation properties are established in (3.39),

$$Q_5 = \int \mathrm{d}r\, j_5^0 \quad , \qquad \tilde{Q}_5 = \int \mathrm{d}r (j_5^0 + 2\hbar \mathscr{C}^0) = Q_5 + 2\hbar W(A) \quad ,$$

$$\mathscr{G}_U \tilde{Q}_5 \mathscr{G}_U^\dagger = \tilde{Q}_5 + 2\hbar n_U \quad . \tag{4.24b}$$

With vector gauge couplings and massless fermions there are also non-singlet axial vector currents j^μ_{5a} which classically are covariantly conserved. (They are not Noether currents since they are not "ordinarily" conserved.) However, the covariant conservation law acquires an anomaly upon quantization,

$$\partial_\mu j^\mu_{5a} + g f_{abc} A^b_\mu j^\mu_{5c} = \frac{\mathrm{i}\hbar}{8\pi^2} \operatorname{tr} \tau^a {}^*\mathscr{F}^{\mu\nu} \mathscr{F}_{\mu\nu} \quad . \tag{4.25}$$

This is the obvious generalization of (4.23); but the anomaly is not a total divergence.

There are also anomalies in source currents for gauge fields when axial vector couplings are present. While there are no purely axial ($\mathrm{i}\gamma^\mu \gamma_5$) non-Abelian theories, chiral couplings ($\frac{1}{2}(1 \pm \mathrm{i}\gamma_5)\gamma^\mu$) can occur if the fermions are massless. For these the functional determinant responds to an infinitesimal gauge transformation in a non-trivial way. We give here the preferred result for a simple group [66, 74]; for direct product groups there is some ambiguity in handling the individual factors; see Exercise 4.2.

$$(D_\mu J^\mu_\pm)_a = \pm \frac{\mathrm{i}\hbar}{24\pi^2} \partial_\mu \epsilon^{\mu\alpha\beta\gamma} \operatorname{tr} \tau^a (\mathscr{A}_\alpha \partial_\beta \mathscr{A}_\gamma + \tfrac{1}{2} \mathscr{A}_\alpha \mathscr{A}_\beta \mathscr{A}_\gamma)$$

$$= \pm \frac{\mathrm{i}\hbar g^2}{24\pi^2} D_{abc} \partial_\mu \epsilon^{\mu\alpha\beta\gamma} (A^b_\alpha \partial_\beta A^c_\gamma + \tfrac{1}{4} g f_{cde} A^b_\alpha A^d_\beta A^e_\gamma) ,$$

$$D_{abc} \equiv \operatorname{tr} \tau^a \tfrac{1}{2} \{\tau^b, \tau^c\} \quad . \tag{4.26}$$

If (4.26) is non-vanishing, the fermionic determinant is not gauge invariant and the gauge theory is rejected. Note that the anomalous divergence is not gauge covariant, which highlights once again that gauge invariance has been lost. In particular (4.26) does not have the same structure as (4.23) with an additional τ^a matrix inside the trace, nor is it of the form (4.25).

One may understand the total derivative structure of the anomaly (4.26) in the following way. Recall from Exercise 2.19 that the conservation of the Noether current j^μ arising from the rigid gauge invariance of the Lagrangian is equivalent to the Yang-Mills field equation and the covariant conservation of the source current J^μ. If the latter is not covariantly conserved, j^μ is not conserved. However, if $D_\mu J^\mu$ can be expressed as a total divergence, then j^μ may be modified and a conserved current may be defined. This indicates that the anomaly (4.26) leads to a breakdown of the local gauge symmetry, but not of the rigid gauge symmetry.

Exercise 4.2. Consider the SU(2) × U(1) group, with a singlet vector potential a_μ, as well as a triplet of SU(2) gauge potentials A^a_μ, coupling chirally to a massless fermion doublet. The determinant is

$$\Delta(A) = \det{}^{1/2}\left(\not\partial + g\,\frac{\sigma^a}{2\mathrm{i}}\,\not A_a + \frac{g'}{2\mathrm{i}}\,\not a\right) \quad . \tag{E4.1}$$

Compute the divergence of the singlet current using (4.26), and observe that it is not SU(2) gauge invariant. Is it U(1) gauge invariant? Compute also the divergence of the triplet current and show that it is neither SU(2) nor U(1) gauge covariant.

Show that one may modify the (unspecified) definition of $-\mathrm{i}\hbar \ln \Delta(A)$ which gives (4.26) by adding a local term, proportional to

$$\int \mathrm{d}x\, \epsilon^{\mu\alpha\beta\gamma}\, \epsilon_{abc}\, a_\mu A^a_\alpha A^b_\beta A^c_\gamma \quad ,$$

so that the singlet current anomalous divergence is now given by an SU(2) × U(1) gauge invariant expression. What is the (modified) divergence of the SU(2) current? Alternatively, can the triplet current divergence be SU(2) × U(1), SU(2) or U(1) gauge covariant? If so, what is the modified divergence of the singlet current?

Exercise 4.3. Derive the commutator algebra of the operators $(D_\mu\, \partial/\delta A_\mu)_a$ and compare with (3.26). By applying this commutator to $\det(\not\partial + \not A)$, find an integrability condition that the anomalous divergence of J^μ_a must satisfy [75]. Verify that (4.26) satisfies this condition, but that (4.25) does not. This condition is called the Wess-Zumino consistency conditon.

Let us explain in detail the reason why anomalies (4.25) and (4.26) differ in form. In the latter $J^\mu_\pm$ is given by the gauge variation of a functional — the determinant for fermions of positive or negative chirality — and as a consequence its divergence satisfies the Wess-Zumino integrability condition, see Exercise 4.3. In the former j^μ_5 is not the variation of anything; the Noether current j^μ_5 does not result from a variation, since nothing couples to it. Thus, its anomalous divergence need not satisfy the Wess-Zumino condition, and one checks that (4.25) does not, see Exercise 4.3. (Statements in the literature that an anomaly "must" satisfy the Wess-Zumino condition [61, 75, 76] are inaccurate; recently, this confusion has been extensively elucidated [77].) On the other hand, since (4.25) arises in the consistent, gauge invariant theory of QCD, it should be gauge invariant, as indeed it is, while the gauge non-invariant formula (4.26) arises in a theory which has lost gauge invariance.

In spite of the quite different physical settings for the two anomalies, (4.25) and (4.26), there is a mathematical relationship between them. Observe the identity

$$\partial_\mu \epsilon^{\mu\alpha\beta\gamma} (\mathscr{A}_\alpha \partial_\beta \mathscr{A}_\gamma + \tfrac{1}{2} \mathscr{A}_\alpha \mathscr{A}_\beta \mathscr{A}_\gamma)$$

$$+ D_\mu \epsilon^{\mu\alpha\beta\gamma} (\mathscr{A}_\alpha \partial_\beta \mathscr{A}_\gamma + \partial_\beta \mathscr{A}_\gamma \mathscr{A}_\alpha + \tfrac{3}{2} \mathscr{A}_\alpha \mathscr{A}_\beta \mathscr{A}_\gamma)$$

$$= \frac{3}{2} {}^*\mathscr{F}^{\mu\nu} \mathscr{F}_{\mu\nu} \tag{4.27}$$

This means that if we add to $J^\mu_\pm$

$$\Delta J^\mu_\pm \equiv \pm \frac{i\hbar}{24\pi^2} \epsilon^{\mu\alpha\beta\gamma} \operatorname{tr} (\mathscr{A}_\alpha \partial_\beta \mathscr{A}_\gamma + \partial_\beta \mathscr{A}_\gamma \mathscr{A}_\alpha + \tfrac{3}{2} \mathscr{A}_\alpha \mathscr{A}_\beta \mathscr{A}_\gamma) \tag{4.28}$$

we shall obtain currents whose covariant divergence is $(\pm i\hbar/16\pi^2) {}^*\mathscr{F}^{\mu\nu}\mathscr{F}_{\mu\nu}$ [78]. Since j^μ_5 is the difference of the right chiral current and the left current we recognize in this way the formula (4.25).

The meaning of this manipulation is the following. Regardless whether a current is gauge source current or a Noether current, the same Feynman diagrams describe its matrix elements. However, different local terms, which are not determined by the diagrams, are appropriate in the two cases, and the above addition $\Delta J^\mu_\pm$ reflects the difference between the local terms contributing to $J^\mu_\pm$ and those in j^μ_5. Moreover, $\Delta J^\mu_\pm$ will not in general be integrable with respect to $\mathscr{A}^\mu$, and that is why j^μ_5 will not be the variation of anything, nor will its divergence satisfy the Wess-Zumino condition.

The quantity D_{abc} in (4.26) must vanish, if chiral couplings are to be gauge invariant [66]. Two cases may be distinguished. It may be that for all representations of the group, $D_{abc} = 0$; these are called "safe" groups and they include SU(2) but no other special unitary groups, all orthogonal groups except SO(6) ≈ SU(4), and all symplectic groups. On the other hand even if the group is not safe, like SU(N), with $N > 2$, it may still be that for some particular representations D_{abc} vanishes [79]. This then gives a useful limitation on the allowed fermion representations.

For an Abelian gauge theory, an axial coupling is classically allowed if the fermions are massless — we are speaking of axial electrodynamics. But again, the axial vector current is not conserved, owing to the three axial current triangle graph [73]; so axial quantum electrodynamics is not gauge invariant.

One believes that the coefficients of the anomalies are not modified by radiative corrections; a theorem proven for the vector Abelian theory — electrodynamics [81]. Although the proof is technical, depending on details of the renormalization procedure, the idea is simple. If one regulates the photons, all divergences of the theory are removed, save those associated with graphs containing fermion loops without photon insertions. But regulating photons does not interfere with chiral symmetry, so chiral anomalies should arise only from lowest-order fermion loops, with no further corrections. (Correspondingly, scaling anomalies associated with the trace of the energy-momentum tensor do possess corrections [10, 12] because any regulator violates scale invariance.) Presumably the result holds in the non-Abelian theory as well, but it may very well be regularization dependent. For example in a supersymmetric theory, one should treat bosons and fermions on equal footing, so regulating only boson lines would be inappropriate [82].

I shall postpone discussing the behaviour of the functional fermion determinant under finite gauge transformations. Suffice it to say here that physically interesting theories, like quantum chromodynamics with vector couplings and unified models with chiral couplings, are invariant against finite gauge transformations as soon as they are infinitesimally invariant.

4.4 Some Physical Consequences of Anomalies

We have seen that in two dimensions the axial anomaly is responsible for generating a vector meson mass. Moreover, in physical four-dimensional theories, the chiral anomaly has many applications [61]. We discuss here the most important ones.

The Glashow-Weinberg-Salam unified theory of electro-weak interactions utilizes chiral SU(2) × U(1) couplings. Since that is not a safe group, one must insure the fermions lie in safe representations. This can be done, provided quarks and leptons balance in number. Thus the requirement that the standard model be anomaly-free leads to the prediction that for every observed lepton there should exist a quark [66]. The prediction has thus far been verified; most recently the discovery of the τ lepton was soon followed by evidence for the "bottom" quark. Now we are anxiously awaiting word about the τ neutrino and the "top" quark.

Once the chiral source currents for the electro-weak gauge fields are conserved, the baryon number current acquires an anomaly [83] [just as in the two-dimensional example the fermion number current is anomalous with gauge-invariant axial-vector interactions; see (4.16)].

$$\partial_\mu j^\mu = \frac{\hbar}{8\pi^2} \operatorname{tr} {}^*\mathscr{F}^{\mu\nu}\mathscr{F}_{\mu\nu} \quad , \tag{4.29}$$

Since the divergence of the baryon number current is proportional to $\operatorname{tr} {}^*\mathscr{F}^{\mu\nu}\mathscr{F}_{\mu\nu}$, whenever that quantity is sizeable in a quantum process one may expect "topological" baryon decay (which should not be confused with "ordinary" baryon decay in Grand Unified Theories). Two mechanisms for topological baryon decay are known. The first involves tunnelling, and in a semi-classical description instantons are the dominant field configurations. For these, $\operatorname{tr} {}^*\mathscr{F}^{\mu\nu}\mathscr{F}_{\mu\nu}$ clearly is sizeable, but the tunnelling rate, being exponentially small, is negligible [84]. However, as mentioned earlier, the Pontryagin index of an 't Hooft-Polyakov monopole is also nonzero [52]; therefore one expects baryon decay in the presence of a monopole. While the magnitude of this process is still controversial, there are arguments that it is large [41]. Nevertheless, practical significance is obscured by the absence of any experimental evidence for monopoles, other than just one reported sighting [39].

Although I have not discussed anomalies in "partially conserved" currents, one physical effect should be mentioned, since historically it opened the subject [65]. The hypothesis of partial conservation of flavor SU(2) axial vector currents (PCAC) implies, in the absence of anomalies, that a massless neutral pion cannot decay into two photons [85]. But the physical pion does decay, with a width of about 7.9 eV. This cannot be accounted for by the finite mass of the physical pion. However, taking into account the anomaly in the axial vector current of the type (4.21), which arises from electromagnetic couplings, one obtains a non-vanishing result [86], that depends on the number of quark colors. Excellent agreement (about 10% too small) with the experimental number is gotten for three colors. (The remaining discrepancy is attributed to finite mass effects of the pion.) Therefore the anomaly provides an experimental determination of the number of colors.

For the final application, let us return to the four-dimensional vectorial Yang-Mills theory, and inquire how the addition of fermions affects the vacuum angle and the semi-classical picture of tunnelling. We discuss first massless Dirac fermions, whose axial vector Noether current is anomalously non-conserved, according to (4.23). For definiteness we consider a SU(2) gauge theory with one fermion doublet, and we return to the Hamiltonian formalism.

From (4.24) one sees that a conserved current does exist, and the time-independent charge $\tilde{Q}_5$ is composed of two pieces; a gauge invariant contribution coming from the fermions, and an anomalous term constructed from gauge potentials, which we recognize to be twice $\hbar W(A)$, defined in (3.33). As mentioned earlier, $\tilde{Q}_5$ is not invariant against homotopically non-trivial, static gauge transformations; rather $\tilde{Q}_5$ shifts by twice the winding number of the gauge function. The commutator algebra of the three operators H, $\tilde{Q}_5$ and $\mathscr{G}_n$ is

$$[H, \tilde{Q}_5] = 0 \quad , \qquad [H, \mathscr{G}_n] = 0, \qquad [\mathscr{G}_n, \tilde{Q}_5] = 2\hbar n \mathscr{G}_n \ . \tag{4.30}$$

The three cannot be simultaneously diagonalized, and gauge invariance requires that physical states be eigenstates of $\mathscr{G}_n$. But this means that $\tilde{Q}_5$ acts as a lowering operator for θ.

$$\exp\left(\frac{i}{\hbar}\,\theta'\tilde{Q}_5\right)\Psi_\theta = \Psi_{\theta-2\theta'} \ . \tag{4.31}$$

Energy eigenvalues of H, which commutes with $\tilde{Q}_5$, can no longer depend on θ; tunnelling is suppressed, and the entire energy band collapses to one level. Physical, gauge and chiral invariant quantities cease to depend on θ. Moreover, chiral symmetry is spontaneously broken because states are not chirally invariant. However, this spontaneous breaking does not derive from energetic stability reasons as in the Goldstone-Nambu mechanism, but rather it occurs because of the axial vector anomaly [46].

The same results may be seen in a functional integral formulation [47]. The generalization of (3.43) to include fermions is (gauge fixing terms suppressed)

$$Z_\theta = \int \mathscr{D}\psi\mathscr{D}\bar{\psi}\mathscr{D}A^a_\mu \exp\left(\frac{i}{\hbar}\int dx\,\mathscr{L}\right) \ , \tag{4.32}$$

where

$$\mathscr{L} = \frac{1}{2g^2} \operatorname{tr} F^{\mu\nu} F_{\mu\nu} - \frac{\hbar\theta}{16\pi^2} \operatorname{tr} {}^*F^{\mu\nu} F_{\mu\nu} + i\hbar \bar{\psi} \gamma^\mu (\partial_\mu + \mathscr{A}_\mu) \psi \quad .$$

When the fermionic integration variable is redefined by a chiral transformation,

$$\psi \to e^{-\theta' \gamma_5} \psi \quad , \qquad \bar{\psi} \to \bar{\psi} e^{-\theta' \gamma_5} \quad , \tag{4.33}$$

it would appear that the functional integral is left invariant. However, the axial anomaly indicates that this is not so; rather Z_θ changes according to

$$Z_\theta \to Z_{\theta - 2\theta'} \quad . \tag{4.34}$$

The detailed reason for non-invariance of the superficially invariant integral in (4.32) has been traced to the singular nature of the fermion measure [87]. Physical quantities cannot be affected by changing integration variables but such changes modify θ, so we must conclude that in the presence of massless fermions the angle is not a physical parameter and can be set to zero.

In a semi-classical treatment, the suppression of tunnelling is recognized after the fermion integration is performed. Then (4.32) leaves

$$Z_\theta = \int \mathscr{D} A^a_\mu \det(\not{\partial} + \not{\mathscr{A}}) \times \exp \frac{i}{\hbar} \int dx \left\{ \frac{1}{2g^2} \operatorname{tr} F^{\mu\nu} F_{\mu\nu} - \frac{\hbar\theta}{16\pi^2} {}^*F^{\mu\nu} F_{\mu\nu} \right\} \quad . \tag{4.35}$$

When this integral is continued to Euclidean space, and dominated by a tunnelling instanton configuration, it vanishes because the Euclidean Dirac operator $(\not{\partial} + \not{\mathscr{A}})$ possesses a zero eigenvalue [88], thus forcing the determinant to vanish. Indeed, the number of zero modes of the Dirac operator is counted by the Pontryagin index; a fact which relates the axial vector anomaly to the topological Atiyah-Singer index theorem [89].

Exercise 4.4. Solve the four-dimensional Euclidean Dirac equation in the instanton field (3.60) and find the expression for the zero mode eigenfunction.

But physical fermions are not massless, and the θ angle in the physical quantum chromodynamical model remains observable. This now presents a problem. Recall that θ is a *CP* violating parameter, but the stringent experimental limits on the neutron's electric dipole moment require that it be practically zero. (More accurately, the limit is $\theta \lesssim 10^{-9}$ [90].) Yet there is no known, physically acceptable principle which insures the vanishing of θ. In particular, setting it to zero *ab initio* would not help for the following reason. Although the origin of fermion masses remains obscure, we expect that spontaneous symmetry breaking is responsible. In that context, the fermion mass matrix would arise in the quantum chromodynamical Lagrangian pointing in an arbitrary *CP* direction, i.e., its form would be $\bar{\psi} M_1 \psi + \bar{\psi} \gamma_5 M_2 \psi$. In order to isolate *CP* violating effects, one needs to remove the term involving M_2. This may be achieved by a chiral redefinition of the Fermi fields, which formally leaves the rest of the Lagrangian invariant, but actually — because of the axial vector anomaly — induces a tr $^*\mathscr{F}^{\mu\nu} \mathscr{F}_{\mu\nu}$ term, giving rise to a vacuum angle. Thus what is needed is a principle that would insure that the "initial" value of θ be precisely cancelled by this chiral redefinition — but such a principle is missing.

We are facing a problem not unlike that of the cosmological constant in gravity theory. The same general principles of invariance and renormalizability which select the kinetic part of the Lagrangian, allow the additional constant. Experimental observation, however, requires it to be zero. But setting the constant to zero initially does not help because spontaneous symmetry breaking gives rise to it anew. Of course one difference is that the cosmological constant modifies the classical theory, while the vacuum angle is a quantum effect.

Fortunately there is also good news for quantum chromodynamics from the θ angle and the associated phenomena. For a long time it appeared that the theory possesses too much symmetry to be phenomenologically acceptable, since it was not realized that $\tilde{Q}_5$ is gauge variant. This symmetry predicts that there would be a particle degenerate with the pion, and no such particle exists [91]. Now we recognize this to be a false prediction. The so-called U(1) problem has dissolved [92]!

5. Quantization Constraints on Physical Parameters

We have seen that quantum mechanics and gauge invariance constrain the structure of a consistent quantum field theory: the theory must be anomaly free, and if there is a possibility of anomalies, fermions must transform