PHYSICS REPORTS (Review Section of Physics Letters) 142, No. 6 (1986) 357–387. North-Holland, Amsterdam

HOW INSTANTONS SOLVE THE U(1) PROBLEM

G. 't HOOFT

Institute for Theoretical Physics, Princetonplein 5, P.O. Box 80.006, 3508 TA Utrecht, The Netherlands

Received 18 April 1986

Contents:

Abstract:

The gauge theory for strong interactions, QCD, has an apparent U(1) symmetry that is not realized in the real world. The violation of the U(1) symmetry can be attributed to a well-known anomaly in the regularization of the theory, which in field configurations called "instantons" can be seen to give rise to interactions that explicitly break the symmetry. A simple polynomial effective Lagrangian describes these effects qualitatively very well. In particular it is seen that no unwanted Goldstone bosons appear and the eta particle owes a large fraction of its mass to instantons. There is no need for field configurations with fractional winding numbers and it is explained how a spurious U(1) symmetry that remains in QCD even after introducing instantons, does not affect these results.

Single orders for this issue

PHYSICS REPORTS (Review Section of Physics Letters) 142, No. 6 (1986) 357–387.

Copies of this issue may be obtained at the price given below. All orders should be sent directly to the Publisher. Orders must be accompanied by check.

Single issue price Dfl.22.00, postage included.

HOW INSTANTONS SOLVE THE U(1) PROBLEM

G. 't HOOFT

Institute for Theoretical Physics, Princetonplein 5, P.O. Box 80.006, 3508 TA Utrecht, The Netherlands

NORTH-HOLLAND - AMSTERDAM

Editorial Note

A review article should not lead to any major controversy even if it often implies that the author takes sides when conflicting results or approaches exist. Yet, when it covers a delicate and topical question, it may happen that some important controversy escapes the editor's attention. The very important and difficult "U(1) problem" was reviewed by G.A. Christos in a Physics Reports article published in 1984.

The present article by G. 't Hooft is meant to be a critical supplement to this former review. The editor is thankful to the present author to thus help to clarify this difficult question for the readers of Physics Reports.

M. Jacob

1. Introduction

In addition to the usual hadronic symmetries the Lagrangian of the prevalent theory of the strong interactions, quantum chromodynamics (QCD) shows a chiral U(1) symmetry which is not realized, or at least badly broken, in the real world [1]. Now although it was soon established that the corresponding current conservation law is formally violated by quantum effects due to the Adler–Bell–Jackiw anomaly [2] it was for some time a mystery how effective U(1) violating interactions could take place to realize this violation, in particular because a less trivial variant of chiral U(1) symmetry still seemed to exist. Indeed, all perturbative calculations showed a persistence of the U(1) invariance.

With the discovery of instantons [3], and the form the Adler–Bell–Jackiw anomaly takes in these nonperturbative field configurations, this so-called U(1) problem was resolved [4, 5]. It was now clear how entire units of axial U(1) charge could appear or disappear into the vacuum without the need of (nearly) massless Goldstone bosons. In a world without instantons the η and η' particles would play the role of Goldstone bosons. Now the instantons provide them with an anomalous contribution to their masses.

In the view of most theorists the above arguments neatly explain why the η particle is considerably heavier than the pions (one must compare m_η^2 with m_π^2), and η' much heavier than the kaons.

Not everyone shares this opinion. In particular Crewther [6] argues that Ward identities can be written down whose solutions would still require either massless Goldstone bosons or gauge field configurations with fractional winding numbers, whereas experimental evidence denies the first and index theorems in QCD disfavor the second. If he were right then QCD would seem to be in serious trouble.

In a recent review article this dissident point of view was defended [7]. The way in which it refers to the present author's work calls for a reaction. At first sight the disagreement seems to be very deep. For instance the sign of the axial charge violation by the instanton is disputed; there are serious disagreements on the form of the effective interaction due to instantons, and the Crewther school insists that chiral U(1) is only spontaneously broken whereas we prefer to call this breaking an explicit one. Now as it turns out after closer study and discussions [8], much of the disagreement (but not all) can be traced back to linguistics and definitions. The aim of this paper is to demonstrate that using quite reasonable definitions of what a "symmetry" is supposed to mean for a theory, the "standard view" is absolutely correct; chiral U(1) is explicitly broken by instantons, and the sign of ΔQ_5 is as given by the anomaly equation; the effective Lagrangian due to instantons can be chosen to be local and polynomial in the mesonic fields, and the η and η' acquire masses due to instantons with integer winding numbers.

In order to make it clear to the reader what we are talking about we first consider a simple model (section 2) which we claim to be the relevant effective theory for the mesons in QCD (even though it is dismissed by ref. [7]). The model shows what the symmetry structure of the vacuum is and how the η particles obtain their masses. It also exhibits a curious periodicity structure with respect to the instanton θ angle, which was also noted in [7] but could not be cast in an easy language because of their refusal to consider models of this sort.

Now does the model of section 2 reflect the symmetry properties of QCD properly? Since refs. [6, 7, 8] express doubt in this respect we show in section 3 how it reflects the exactly defined operators and Green's functions of the exact theory. The calculation of the η-mass is redone, but now in terms of QCD parameters. We clearly do not pretend to "solve" QCD, so certain assumptions have to be made. The most important of these, quite consistent with all we know about QCD and the real world, is

$$\langle \bar{q}q \rangle = F \neq 0 . \tag{1.1}$$

Now refs. [6, 7] claim that in that case one needs gauge field configurations with fractional winding numbers. Our model calculations will clearly show that this is not the case.

How can it be then if our calculations appear to be theoretically sound and in close agreement with experiment, that they seem to contradict so-called "anomalous Ward identities"? To answer this question we were faced with unraveling some problems of communication. The current-algebraic methods of refs. [6, 7] were mainly developed before the rise of gauge theories but are subsequently applied to gauge-noninvariant sectors of Hilbert space. Their language is quite different from the one used in many papers on gauge theories [4] and due to incorrect "translations" several results of the present author were misquoted in [7]. After making the necessary corrections we try to analyze, in our own language, where the problem lies.

There are two classes of identities that one can write down for Green functions. One is the class of identities that follow from exactly preserved global or local symmetries. Local symmetries must always be exact symmetries. From those symmetries we get Ward identities [9] (in the Abelian case), or more generally Ward–Slavnov–Taylor [10] identities, which also follow from the (exact) Becchi–Rouet–Stora [11] global invariance.

On the other hand we have identities which follow from applying field transformations which may have the form of gauge transformations but which do not leave the Lagrangian (or, more precisely, the entire theory) invariant. These transformations are sometimes called Bell–Treiman transformations [12] in the literature, but "Veltman transformations" would be more appropriate [13]. The identities one gets reflect to some extent the dynamics of the theory and form a subclass of its Dyson–Schwinger equations. Thus when refs. [6, 7] perform chiral rotations of the fermionic fields, which do not leave θ invariant, they are performing a Veltman transformation, and their so-called "anomalous Ward identities" fall in this second class of equations among Green functions.

It is in the second derivative with respect to θ that refs. [6, 7] claim to get contradiction with our model calculations [8]: the second derivative of an insertion of the form

$$\mathrm{i}\theta F\tilde{F} , \tag{1.2}$$

in the Lagrangian vanishes, whereas the simple "effective field theory" requires an insertion of the form

$$\mathrm{e}^{\mathrm{i}\theta} \det(q_{\mathrm{L}}\bar{q}_{\mathrm{R}}) + \mathrm{h.c.} , \tag{1.3}$$

of which the second derivative produces the η mass. So they claim that in the real theory the η mass cannot be explained that simply.

In section 5 we explain why, in the modern formalism this problem does not arise at all. (1.2) should not be confused with ordinary Lagrange insertions and after resummation correctly reproduces (1.3). The canonical methods are not allowed if one tries first to quantize the gauge fields A_μ and only afterwards the fermion fields. The fermions have to be quantized and integrated out first. The phenomenon of "variable numbers of canonical fermionic variables" resolves the dilemma. We conclude, that the η mass is what it should be and there is no U(1) problem.

There is a further linguistic disagreement on whether the U(1) breaking of QCD should be called an explicit or a spontaneous symmetry breaking. In the effective Lagrangian model the symmetry is clearly broken explicitly. Several authors however refer to the θ angle as a property of the vacuum [5] in QCD. Of course as long as one agrees on the physical effects one is free to use whatever terminology seems appropriate. We merely point out that all physical consequences of the instanton effects in QCD (in particular the absence of a physical Goldstone boson) coincide with the ones of an explicit symmetry breaking. Our θ angle is as much a constant of Nature as any other physical parameter, to be compared for instance with the electron mass term which breaks the electron's chiral invariance.

Only if one adds nonphysical sectors to Hilbert space one may obtain an alternative description of the θ angle as a parameter induced by boundary effects producing a spontaneous symmetry breakdown. However, *any* explicit symmetry breaking can be turned into "spontaneous" symmetry breaking by artificially enlarging the Hilbert space. One gets no physically observable Goldstone boson however, so, the most convenient place to draw the dividing line between spontaneous and explicit global symmetry breaking is between the presence or absence of a Goldstone boson.

We explain this situation in section 6, where we show that any nonphysical symmetry can be forced upon a theory this way. In section 7 we show how the spurious U(1) symmetry that is used as a starting point in [7], actually belongs to this class. Section 8 shows that if one takes into account the enlarged Hilbert space with the variable θ angles then the effective theory of section 2 neatly obeys the so-called anomalous Ward identities. The effective model shows the vacuum structure so clearly that all problems with "fractional winding numbers" are removed.

The decay amplitude $\eta \rightarrow 3\pi$ posed problems similar to that of the η mass. As explained correctly in [7] this problem is resolved as soon as the U(1) breaking is understood so no further discussion of this decay is necessary. It fits well with theory.

Appendix A is a comment concerning the sign of axial charge violation under various boundary conditions.

Appendix B discusses the instanton-induced amplitude. The contribution from "small" instantons can be computed precisely but the infrared cutoff is uncertain. (Rough) Estimates of the amplitude in a simple color SU(2) theory give quite large values, which confirms that instantons may affect the symmetry structure of QCD sufficiently strongly such as to explain the known features of hadrons.

2. A simple model

Before really touching upon some of the more subtle aspects of the "U(1) problem" we first construct a simple "effective Lagrangian" model. Whether or not this model truly reflects the symmetry properties of QCD (which we do claim to be the case) is left to be discussed in the following sections. For simplicity the model of this section will be discussed only in the tree-approximation.

To be explicit we take the number of quark flavors to be two. Generalization towards any numbers of flavors (two and three are the most relevant numbers to be compared with the situation in the real world) will be completely straightforward at all stages in this section*. So we start with the "unbroken model" having global invariance of the form

* See however the remark following eq. (2.24).

$$\mathrm{U}(L)_\mathrm{L} \otimes \mathrm{U}(L)_\mathrm{R}\,, \qquad \text{with } L = 2\,, \tag{2.1}$$

where the subscripts L and R refer to left and right, respectively. We consider complex meson fields with the quantum numbers of the quark–antiquark composite operator $\bar{q}_{\mathrm{R}j} q_{\mathrm{L}i}$. They transform under (2.1) as:

$$\phi'_{ij} = U^\mathrm{L}_{ik} \phi_{kl} U^{\mathrm{R}\dagger}_{lj}\,, \tag{2.2}$$

to be written simply as

$$\phi' = U^\mathrm{L} \phi U^{\mathrm{R}\dagger}\,. \tag{2.3}$$

Since we have no hermiticity condition on ϕ, there are eight physical particles σ, η, π_a and α_a ($a = 1, 2, 3$):

$$\phi = \tfrac{1}{2}(\sigma + \mathrm{i}\eta) + \tfrac{1}{2}(\boldsymbol{\alpha} + \mathrm{i}\boldsymbol{\pi}) \cdot \boldsymbol{\tau}\,, \tag{2.4}$$

where $\tau^{1,2,3}$ are the Pauli matrices. We take as our Lagrangian:

$$\mathscr{L} = -\mathrm{Tr}\, \partial_\mu \phi \partial_\mu \phi^\dagger - V(\phi)\,. \tag{2.5}$$

A potential V_0 invariant under (2.1) is

$$V_0(\phi) = -\mu^2\, \mathrm{Tr}\, \phi\phi^\dagger + \tfrac{1}{2}(\lambda_1 - \lambda_2)(\mathrm{Tr}\, \phi\phi^\dagger)^2 + \tfrac{1}{2}\lambda_2\, \mathrm{Tr}(\phi\phi^\dagger)^2 \tag{2.6}$$

$$= -\frac{\mu^2}{2}(\sigma^2 + \eta^2 + \boldsymbol{\alpha}^2 + \boldsymbol{\pi}^2) + \frac{\lambda_1}{8}(\sigma^2 + \eta^2 + \boldsymbol{\alpha}^2 + \boldsymbol{\pi}^2)^2 + \frac{\lambda_2}{2}((\sigma\boldsymbol{\alpha} + \eta\boldsymbol{\pi})^2 + (\boldsymbol{\alpha} \wedge \boldsymbol{\pi})^2)\,. \tag{2.7}$$

Assuming, as usual

$$\langle \sigma \rangle = f\,, \qquad \sigma = f + s\,, \tag{2.8}$$

we get, by taking the extremum of (2.7),

$$f^2 = 2\mu^2/\lambda_1\,; \tag{2.9}$$

$$V_0 = \frac{\lambda_1}{2}(fs + \tfrac{1}{2}(s^2 + \eta^2 + \boldsymbol{\alpha}^2 + \boldsymbol{\pi}^2))^2 + \frac{\lambda_2}{2}((f\boldsymbol{\alpha} + s\boldsymbol{\alpha} + \eta\boldsymbol{\pi})^2 + (\boldsymbol{\alpha} \wedge \boldsymbol{\pi})^2)\,, \tag{2.10}$$

from which we read off:

$$m_\mathrm{s}^2 = \lambda_1 f^2 = 2\mu^2\,, \qquad m_\eta^2 = 0\,, \qquad m_\alpha^2 = \lambda_2 f^2\,, \qquad m_\pi^2 = 0\,. \tag{2.11}$$

There are four Goldstone bosons, as expected from the U(2)⊗U(2) invariance, broken down to U(2) by (2.8).

We now consider two less symmetric additional terms in V:

$$V_m = U_m + U_m^* ;$$

$$U_m = \text{“} m\, e^{i\chi} \bar{q}_R q_L \text{”} = \tfrac{1}{4} m\, e^{i\chi}\, \mathrm{Tr}\, \phi = \tfrac{1}{2} m\, e^{i\chi} (\sigma + i\eta) , \tag{2.12}$$

and:

$$V_a = U_a + U_a^* ;$$

$$U_a = \text{“} \kappa\, e^{i\theta} \det(\bar{q}_R q_L) \text{”} = \kappa\, e^{i\theta} \det \phi$$

$$= \kappa\, e^{i\theta} ((\sigma + i\eta)^2 - (\boldsymbol{\alpha} + i\boldsymbol{\pi})^2) . \tag{2.13}$$

Here, m, χ, κ and θ are all free parameters. The terms between quotation marks are there just to show the algebraic structure, up to renormalization constants. Notice that V_0 still has the U(1) invariance

$$\phi \to e^{i\omega} \phi ,$$

or

$$(\sigma + i\eta) \to e^{i\omega} (\sigma + i\eta) , \qquad (\boldsymbol{\alpha} + i\boldsymbol{\pi}) \to e^{i\omega} (\boldsymbol{\alpha} + i\boldsymbol{\pi}) . \tag{2.14}$$

Therefore we are free to rotate

$$\chi \to \chi + \omega , \qquad \theta \to \theta + 2\omega . \tag{2.15}$$

(Here the 2 would be replaced by L in a theory with L flavors.)

Consider first the theory with $\kappa = 0$;

$$V = V_0 + V_m . \tag{2.16}$$

Then we can choose $\omega = \pi - \chi$, and

$$V_m = -m\sigma . \tag{2.17}$$

So χ is unphysical. Equation (2.9) is replaced by

$$f^2 = 2\mu^2/\lambda_1 + 2m/\lambda_1 f . \tag{2.18}$$

Consequently, in (2.11) we get

$$m_\pi^2 = m_\eta^2 = m/f . \tag{2.19}$$

Note that with the sign choice of (2.17), f must be the positive solution of (2.18).

Now take the other case, namely $m = 0$;

$$V = V_0 + V_a . \tag{2.20}$$

It is now convenient to choose

$$\omega = \tfrac{1}{2}(\pi - \theta) . \tag{2.21}$$

So here the angle θ is unphysical.

$$V_a = -2\kappa(\sigma^2 + \boldsymbol{\pi}^2 - \eta^2 - \boldsymbol{\alpha}^2) . \tag{2.22}$$

$$f^2 = 2\mu^2/\lambda_1 + 8\kappa/\lambda_1 . \tag{2.23}$$

The masses of the light particles become

$$m_\pi^2 = 0 ; \qquad m_\eta^2 = 8\kappa . \tag{2.24}$$

So the κ term contributes directly to the η mass and not to the pion mass. In case of more than two flavors the determinant in (2.13) will contain higher powers of ϕ, and extra factors f will occur in (2.24). But the mechanism generating a mass for the η' will not really be different from that of the η.

It is interesting to study the case when both m and κ are unequal to zero. We then cannot rotate both χ and θ independently and one of the two angles is physical. Since obviously the model Lagrangian is periodic with period 2π in θ, its physical consequences will be periodic in χ with period π, because of the invariance (2.15). In the general case we now also expect a nonvanishing value for

$$\langle \eta \rangle = g . \tag{2.25}$$

So we write $\eta = g + h$, where g is a c-number and h a field. The conditions for f and g are:

$$\begin{aligned} &\tfrac{1}{2}\lambda_1 f(f^2 + g^2) - \mu^2 f + m\cos\chi + 4\kappa f\cos\theta - 4\kappa g\sin\theta = 0 ; \\ &\tfrac{1}{2}\lambda_1 g(f^2 + g^2) - \mu^2 g - m\sin\chi - 4\kappa g\cos\theta - 4\kappa f\sin\theta = 0 . \end{aligned} \tag{2.26}$$

It will be convenient however to choose ω in (2.15) such that $g = 0$. According to (2.26) one then must have

$$m\sin\chi + 4\kappa f\sin\theta = 0 . \tag{2.27}$$

Now we find

$$m_\pi^2 = -\frac{m\cos\chi}{f} , \tag{2.28}$$

$$m_\eta^2 = -\frac{m\cos\chi}{f} - 8\kappa\cos\theta , \tag{2.29}$$

apart from an η–s mixing. The effect of this mixing however goes proportional to

$$16\kappa^2 \sin^2 \theta = m^2 \sin^2 \chi \tag{2.30}$$

and therefore is of higher order both in κ and m.

Consider now the function

$$F(\omega) = m \cos(\chi + \omega) + 2\kappa f \cos(\theta + 2\omega)\,. \tag{2.31}$$

Then (2.27) requires ω to be a solution of

$$F'(\omega) = 0 \tag{2.32}$$

(after which we insert the replacement (2.15)), and (2.29) implies:

$$f m_\eta^2 = F''(\omega) > 0\,. \tag{2.33}$$

So, ω must be chosen such that (2.31) takes its minimum value. Furthermore, the replacement $\omega \to \omega + \pi$ switches the sign of the first but not of the second term in $F(\omega)$, so, if ω is chosen to be the *absolute* minimum of (2.31), then indeed also

$$m_\pi^2 > 0\,, \tag{2.34}$$

except when there are two minima. In that case there is a phase transition with long range order (due to the massless pion) at the transition point.

This phase transition is at

$$\begin{aligned} &\chi = \pi/2\,, \\ &\sin\theta = -m/4\kappa f\,, \\ &\cos\theta < 0\,, \end{aligned} \tag{2.35}$$

or, if a rotation (2.15) is performed:

$$\begin{aligned} &\chi = 0\,, \\ &\sin\theta = m/4\kappa f\,, \\ &0 < \theta < \pi/2\,. \end{aligned} \tag{2.36}$$

A further critical point may occur at

$$\begin{aligned} &\chi = 0 \\ &\theta = \pi/2 \\ &m = 4\kappa f \end{aligned} \tag{2.37}$$

where also m_η vanishes. Of course these values of the parameters are not believed to be close to the ones describing real mesons. Phase transitions of this sort were indeed also described in ref. [7]. But because no specific model such as ours was considered, the periodicity structure in χ and θ was discussed in a somewhat untransparent way. Although obviously in our model we have a periodicity in θ with period 2π, ref. [7] suspected a period 4π. Indeed if ω in (2.15) shifts by an amount π, then $m\,e^{i\chi} \to -m\,e^{i\chi}$ and one might end up in an unstable analytic extension of the theory. Clearly the properties of the minimum of the potential $F(\omega)$, eq. (2.31), can only have period 2π in θ.

3. QCD

If Quantum Chromodynamics with L quark flavors, were to have an approximate $U(L)_L \times U(L)_R$ symmetry only broken by quark mass terms, the model of the previous section with $V = V_0 + V_m$ could then conveniently describe the qualitative features of mesons, but with η and π almost degenerate. (In the case $L > 2$ one merely has to substitute the 2×2 matrix ϕ by an $L \times L$ matrix.) That would be the case if somehow the effects of instantons could be suppressed.

Let us now consider instantons and write in a shorthand notation the functional integral I for a certain mesonic amplitude in QCD. For the ease of the discussion we assume all integrations to be in Euclidean space-time:

$$I = \int \mathrm{D}A \int \mathrm{D}\psi\, \mathrm{D}\bar{\psi}\, \exp[S_A + S_{A,\psi} + \mathrm{i}\theta F\tilde{F} - J\bar{\psi}\psi]\,, \tag{3.1}$$

with

$$\mathrm{D}\psi\, \mathrm{D}\bar{\psi} = \prod \mathrm{d}^2\psi_\mathrm{L}(x)\, \mathrm{d}^2\psi_\mathrm{R}(x)\, \mathrm{d}^2\bar{\psi}_\mathrm{L}(x)\, \mathrm{d}^2\bar{\psi}_\mathrm{R}(x)\,, \tag{3.2}$$

$$S_A = \int \mathrm{d}x(-\tfrac{1}{4}F^2_{\mu\nu}(A))\,,$$

$$S_{A,\psi} = -\bar{\psi}_i(\gamma_\mu D^{(A)}_\mu + m_i)\psi_i\,. \tag{3.3}$$

$$F\tilde{F} = \int \mathrm{d}x\, (g^2 F_{\mu\nu}\, \varepsilon_{\mu\nu\alpha\beta} F_{\alpha\beta}/64\pi^2)\,, \tag{3.4}$$

$$J\bar{\psi}\psi = J_\pi(\bar{\psi}_\mathrm{L}\tau\psi_\mathrm{R} - \bar{\psi}_\mathrm{R}\tau\psi_\mathrm{L}) + J_\eta(\bar{\psi}_\mathrm{L}\psi_\mathrm{R} - \bar{\psi}_\mathrm{R}\psi_\mathrm{L}) + \cdots \tag{3.5}$$

As we will see shortly, it is crucial that the integration over ψ is done first and the one over A afterwards.

Since we have no way of solving the theory exactly certain simplifying assumptions must be made. We now claim that the assumptions to be formulated next will in no way interfere with the known symmetry properties of the low-energy theory. Discussion of this claim will be postponed to sections 4–7.

An (anti-)instanton is a field configuration of the A fields with the property

$$\int_{\Delta V} F^a_{\mu\nu} F^a_{\alpha\beta} \varepsilon_{\mu\nu\alpha\beta}\, \mathrm{d}^4x = (-)64\pi^2/g^2\,, \tag{3.6}$$

where ΔV is the volume of a space-time region. Outside ΔV we have essentially $|F_{\mu\nu}| = 0$ but we cannot have $|A| = 0$ there because then (3.6) would vanish as the integrand is a total derivative.

The assumption we make is that the A integral can be split into an integral over instanton-locations and an integral over perturbative fluctuations around those instantons. We do this as follows. Let us divide space-time into four-dimensional boxes with volumes ΔV of the order of $1(\mathrm{fm})^4$. Each box may or may not contain one instanton or one anti-instanton. (There could be more than one instanton or anti-instanton in a single box, but we choose our boxes so small that such multi-instantons in one box become statistically insignificant.) The essential point is that since an instanton in a box ΔV will do nothing but gauge-rotate any of the fields outside ΔV, the instanton-numbers in each box are independent variables. Notice that at this point we do not require these twisted field configurations in the boxes to be exact solutions to the classical field configurations. This is why we have no difficulties confining each instanton to be completely inside one box, with only gauge rotations of the vacuum outside.

Let us then write

$$A = A_{\mathrm{inst}} + \delta A \tag{3.7}$$

where A_{inst} is due to the instantons only, then the integral over δA will essentially commute with the integral over the instantons. The δA integral is assumed to be responsible for the strong binding between the quarks. The confinement problem is *not* solved this way but is not relevant here since we decided to concentrate on low-energy phenomena only.

Note that the integration over A_{inst} is more than a summation over total winding number ν. Rather, if we write

$$\nu = \nu_+ - \nu_- \tag{3.8}$$

then the integral over A_{inst} closely corresponds to integration over the locations of the ν_+ instantons and the ν_- anti-instantons. It is important that we restrict ourselves to instantons with compact support (namely, limited to the confines of the box ΔV in which they belong). A larger instanton, if it occurs, should be represented as a small one in one of the boxes, with in addition a tail that is taken care of by integration over δA. "Very large" instantons are irrelevant because they would be superimposed by small ones. In short, in eq. (3.7), A_{inst} is defined to be a smooth field configuration that accounts for all winding numbers inside the boxes, and δA is defined to contribute to $\int F\tilde{F}$ by less than one unit in each box.

Now consider an isolated instanton located within one of our boxes, located at $x = x_1$. What is discussed at length in the literature is the fact that the ψ integration is now affected by the presence of a zero mode solution of the Dirac equation. If there were no other anti-instantons and no source term J then the fermionic integral, being proportional to the determinant of the operator $\gamma_\mu(\partial_\mu + \mathrm{i}gA_\mu)$, would vanish because of this one zero eigenvalue. If we do add the source term $J\bar{\psi}\psi$ the integral need not vanish. In ref. [4] it was derived that the instanton exactly acts as if it would contain a source for every fermionic flavor. Thus with one instanton located at $x = x_1$ the fermionic integral

$$\int \mathrm{D}\psi\, \mathrm{D}\bar{\psi}\, [\exp(S_{A,\psi} + J\bar{\psi}\psi)] \tag{3.9}$$

has the *same effect* as the integral

$$\kappa \int \mathrm{D}\psi\, \mathrm{D}\bar{\psi}\, [\exp(S_{0,\psi} + J\bar{\psi}\psi)] \cdot \det(\bar{\psi}_{\mathrm{R}}(x_1)\, \psi_{\mathrm{L}}(x_1))\,, \tag{3.10}$$

where κ may be computed from all one-loop corrections [4]. Indeed it was shown that the zero eigenmodes for all flavors which extend beyond the volume ΔV conveniently reproduce the fermionic propagators connecting x_1 with the sources J. The fact that (3.9) does have the same quantum selection properties as (3.10) can also be argued by realizing that a gauge-invariant regulator for the fermions had to be introduced, and instead of the *lowest* eigenmodes one could have concentrated on the much more localized *highest* fermionic states. The correctly regularized fermionic integral contains a mismatch by one unit for each flavor between the total number of left handed and right handed fermionic degrees of freedom. Since this happens both for the fermions and the antifermions $\bar{\psi}$ the determinant in (3.10) consists of products of L fermionic and L antifermionic fields.

Since we do require that the $\mathrm{SU}(L)_{\mathrm{L}} \otimes \mathrm{SU}(L)_{\mathrm{R}}$ is kept unharmed by the instantons, the determinant is at first sight the only allowed choice for (3.10) but, actually, if one does not suppress the color and spin indices, one can write down more expressions with the required symmetry properties.

Next consider ν_+ instantons, located at $x = x_i$. Following a declustering assumption which, at least to the present author's taste, is quite natural and does not require much discussion, we may assume these to act on the fermionic integrations as

$$\kappa^{\nu_+} \int \mathrm{D}\psi\, \mathrm{D}\bar{\psi}\, \mathrm{e}^{S_\psi} \prod_{i=1}^{\nu+} \det(\psi_{\mathrm{L}}(x_i)\, \bar{\psi}_{\mathrm{R}}(x_i))\,. \tag{3.11}$$

Let us add the θ dependence and integrate over the instanton locations x_i:

$$\frac{1}{\nu_+!}\, \mathrm{e}^{\mathrm{i}\theta\nu_+} \prod_i \mathrm{d}^4 x_i\,. \tag{3.12}$$

The denominator $\nu_+!$ is due to exchange symmetry of the instantons.

We now extend our declustering assumption to the anti-instantons as well. This assumption was vigorously attacked in [6–8]. Indeed one might criticize it, for instance by suggesting that "merons" play a more crucial role [14]. We insist however that the assumption in no way interferes with the symmetry properties of our model. We will see in sections 7 and 8 that the anomalous Ward identities will be exactly satisfied by our model. To avoid confusion let us also stress that our declustering assumptions refer to the QCD part of the metric only, *not* to the contributions of the fermions which we denote explicitly. So there is no disagreement at all with the findings of ref. [15]. Indeed, our approach here is closely analogous to theirs.

Thus, consider ν_- anti-instantons. The complete instanton contribution to the functional integral is

$$\sum_{\nu_+=0}^{\infty} \sum_{\nu_-=0}^{\infty} \frac{\kappa^{\nu_+ + \nu_-}}{\nu_+!\,\nu_-!}\, \mathrm{e}^{\mathrm{i}\theta(\nu_+ - \nu_-)} \Big(\int \mathrm{d}^4x \det(\psi_{\mathrm{L}}(x)\bar{\psi}_{\mathrm{L}}(x))\Big)^{\nu_+} \Big(\int \mathrm{d}^4x \det(\psi_{\mathrm{R}}(x)\bar{\psi}_{\mathrm{L}}(x))\Big)^{\nu_-}\,. \tag{3.13}$$

The summations are now easy to carry out:

$$(3.13) = \exp \int \mathrm{d}^4x [\kappa\, \mathrm{e}^{\mathrm{i}\theta} \det \psi_{\mathrm{L}}(x)\bar{\psi}_{\mathrm{R}}(x) + \kappa\, \mathrm{e}^{-\mathrm{i}\theta} \det \psi_{\mathrm{R}}(x)\bar{\psi}_{\mathrm{L}}(x)]\,, \tag{3.14}$$

which is precisely the effective interaction V_{a} of eq. (2.13). The remaining integrals over the fermionic fields ψ and the perturbative fields δA may well result in the effective Lagrangian model of section 2. Notice that, *before* we interchanged the $A_{\rm inst}$ and ψ, $\bar\psi$ integrations, we have made the substitution (3.10). This will be crucial for our later discussions. Once the substitution (3.10) has been made, the (A-field-dependent) extra fermionic degrees of freedom have been taken care of, and only then one is allowed to interchange the A and the ψ integrations. This is how (2.13) follows from (3.14).

4. Symmetries and currents

Let us split the generators $\Lambda_{\rm L}$ and $\Lambda_{\rm R}$ for the $\mathrm{U}(L)_{\rm L}\times \mathrm{U}(L)_{\rm R}$ transformations into scalar ones, Λ^a and Λ^0, and pseudoscalar ones, Λ_5^a and Λ_5^0. The infinitesimal transformation rules for the various fields considered thus far are:

$$\delta\psi_{\rm L} = -\tfrac{1}{2}\mathrm{i}\tau^a(\Lambda^a + \Lambda_5^a)\psi_{\rm L} + \mathrm{i}(\Lambda^0 + \Lambda_5^0)\psi_{\rm L}\,, \tag{4.1}$$

$$\delta\psi_{\rm R} = -\tfrac{1}{2}\mathrm{i}\tau^a(\Lambda^a - \Lambda_5^a)\psi_{\rm R} + \mathrm{i}(\Lambda^0 - \Lambda_5^0)\psi_{\rm R}\,, \tag{4.2}$$

$$\delta\phi = -\tfrac{1}{2}\mathrm{i}\Lambda^a[\tau^a, \phi] - \tfrac{1}{2}\mathrm{i}\Lambda_5^a\{\tau^a, \phi\}_+ + 2\mathrm{i}\Lambda_5^0\phi\,, \tag{4.3}$$

$$\delta\sigma = \Lambda_5^a\pi_a - 2\Lambda_5^0\eta\,, \tag{4.4}$$

$$\delta\eta = -\Lambda_5^a\alpha_a + 2\Lambda_5^0\sigma\,, \tag{4.5}$$

$$\delta\pi_a = \varepsilon_{abc}\Lambda^b\pi_c - \Lambda_5^a\sigma + 2\Lambda_5^0\alpha_a\,, \tag{4.6}$$

$$\delta\alpha_a = \varepsilon_{abc}\Lambda^b\alpha_c + \Lambda_5^a\eta - 2\Lambda_5^0\pi_a\,, \tag{4.7}$$

$$\delta\det\phi = 4\mathrm{i}\Lambda_5^0\det\phi\,. \tag{4.8}$$

In a theory with L flavors the factor 4 in eq. (4.8) must be replaced by $2L$. In a classical field theory the currents are most easily obtained by considering transformations (4.1)–(4.8) with space-time dependent $\Lambda_i(x)$. Their effect on the total action can be written as

$$\delta S = \int \mathrm{d}^4x(-F_i\Lambda_i(x) - J^i_\mu\partial_\mu\Lambda_i(x)) \tag{4.9}$$

(here $i = 1, \dots, 8$).
Since according to the equations of motion $\delta S = 0$ for all choices of $\Lambda_i(x)$, one has

$$\partial_\mu J^i_\mu(x) = F_i(x)\,. \tag{4.10}$$

A Lagrangian which gives invariance under the space-time independent Λ_i must have $F_i = 0$, so that the current J^i_μ is conserved.

We are now mainly concerned about the current $J_{\mu 5}$ associated with Λ_5^0. The QCD Lagrangian (3.1) produces the current

$$J_{\mu 5} = \mathrm{i}\bar{\psi}\gamma_\mu \gamma_5 \psi\,, \tag{4.11}$$

and, prior to quantization:

$$\partial_\mu J_{\mu 5} = 2\mathrm{i}m\bar{\psi}\gamma_5\psi\,. \tag{4.12}$$

As is well known, however, eq. (4.12) does not survive renormalization. Renormalization cannot be performed in a chirally invariant way and therefore the symmetry cannot be maintained, unless we would be prepared to violate the local color gauge-invariance. But violation of color gauge invariance would cause violation of unitarity, so, in a correctly quantized theory, (4.12) breaks down. A diagrammatic analysis [2] shows that, at least to all orders of the perturbation expansion, one gets

$$\partial_\mu J_{\mu 5} = 2\mathrm{i}m\bar{\psi}\gamma_5\psi - \frac{\mathrm{i}Lg^2}{16\pi^2} F^a_{\mu\nu}\tilde{F}^a_{\mu\nu} \tag{4.13}$$

with

$$\tilde{F}^a_{\mu\nu} = \tfrac{1}{2}\varepsilon_{\mu\nu\alpha\beta}F^a_{\alpha\beta}\,. \tag{4.14}$$

We read off that, if we may ignore the mass term, then in a space-time volume V with ν_+ instantons and ν_- anti-instantons

$$\int_{\mathrm{V}} \mathrm{d}^4x\, \partial_\mu J_{\mu 5} = -2\mathrm{i}L(\nu_+ - \nu_-)\,. \tag{4.15}$$

Here the factor i is an artefact of Euclidean space. Defining the charge Q_5 in a 3-volume V_3 by

$$Q_5 = \int_{\mathrm{V}_3} J_0\,\mathrm{d}^3\boldsymbol{x} = \mathrm{i}\int_{\mathrm{V}_3} J_4\mathrm{d}^3\boldsymbol{x} = Q_\mathrm{R} - Q_\mathrm{L}\,, \tag{4.16}$$

each instanton causes a transition*

$$\Delta Q_5 = 2L\,. \tag{4.17}$$

This is called the "naive" equation in ref. [7]. Since we were working in a finite space-time volume V the nature of the "vacuum" has not yet entered into the discussion. Remarks on the language used here and in ref. [7] are postponed to appendix A.

Now let us write the corresponding equation in our effective Lagrangian model. Here,

$$J_{\mu 5} = 2\mathrm{i}\,\mathrm{Tr}\{(\partial_\mu\phi^*)\phi - \phi^*\partial_\mu\phi\} \tag{4.18}$$

and

* Apart from the disputed sign there are also differences in sign conventions with ref. [7].

$$\partial_\mu J_{\mu 5} = -2m(\eta \cos\chi + \sigma \sin\chi) + 16\kappa(\boldsymbol{\alpha}\cdot\boldsymbol{\pi} - \sigma\eta)\cos\theta + 8\kappa(\eta^2 + \boldsymbol{\alpha}^2 - \sigma^2 - \boldsymbol{\pi}^2)\sin\theta\,. \quad (4.19)$$

Before comparing this with eq. (4.13) of the QCD theory let us chirally rotate over an angle $\frac{1}{2}\chi$. The mass term in the original Lagrangian then becomes

$$-\bar{\psi}m\psi\cos\chi - \mathrm{i}\bar{\psi}m\gamma_5\psi\sin\chi\,, \quad (4.20)$$

and then (4.13) becomes

$$\partial_\mu J_{\mu 5} = 2\mathrm{i}m\bar{\psi}\gamma_5\psi\cos\chi - 2m\bar{\psi}\psi\sin\chi - \frac{\mathrm{i}Lg^2}{16\pi^2}F^a_{\mu\nu}\tilde{F}^a_{\mu\nu}\,. \quad (4.21)$$

Therefore the first term in (4.19) can neatly be matched with the first terms of (4.21).

An issue raised in refs. [7, 8] is that there is an apparent discrepancy if we try to identify the last terms of (4.19) with the last term of (4.21). The last term of (4.21) contains the color fields only and there is absolutely no θ dependence here. But the last terms of (4.19) do show a crucial θ dependence. It is essentially

$$8\kappa\,\mathrm{Im}(\mathrm{e}^{\mathrm{i}\theta}\det\phi)\,. \quad (4.22)$$

Where did the θ dependence come from?

One way of arguing would be that the θ dependence of (4.22) is obvious. Chiral transformations are described by eq. (2.15) and any symmetry breaking term in a Lagrangian can obviously not be invariant at the same time. So the θ dependence of (4.22) is as it has to be. The symmetry breaking in QCD is not visible in its Lagrangian but is due to the θ dependence of the regularization procedure.

However, although this argument may explain why (4.19) shows a θ dependence and (4.21) does not, it does not explain why nevertheless these two theories can describe the same system. This is (partly) what the dispute is about. We claim that one can identify in the effective theory

$$\frac{128\pi^2\kappa}{Lg^2}\,\mathrm{Im}(\mathrm{e}^{\mathrm{i}\theta}\det\phi) = -\mathrm{i}F^a_{\mu\nu}\tilde{F}^a_{\mu\nu}\,, \quad (4.23)$$

so that, if $\theta \simeq 0$, one may identify $F\tilde{F}$ with the η field. At the same time we would also like to put

$$\tilde{\phi} = \bar{q}_\mathrm{R} q_\mathrm{L}\,, \quad (4.24)$$

but this θ phase seems to be in disagreement with the canonical quantization procedure if A^a_μ, q_R, q_L and ϕ were to be considered as independent canonical variables. Another way of formulating this problem is that the right hand side of (4.23) seems to commute with the chiral charge operator Q_5 while the left hand side does not.

Notice that if we could somehow suppress instantons essentially $F\tilde{F}$ would vanish. The left hand side of (4.23) would vanish also, because $\kappa \to 0$. This suggests one simple answer to our problem: equation (4.23) violates axial charge conservation, but that is to be expected in a theory where axial charge is not conserved. Unfortunately some physicists insist in considering the U(1) violation by instantons as being "spontaneous" rather than explicit and therefore they rejected this simple answer. A rather curious attempt to bypass the problem was described in [7]. They first propose to replace our V_a of eq. (2.13) by

$$V'_a = \kappa\, \mathrm{Tr}(\log(\phi/\phi^\dagger))^2 . \tag{4.25}$$

But this also does not commute with the axial charge operator and furthermore the logarithm is not single-valued so (4.25) makes no sense at all. So then they propose

$$\int V''_a = \kappa \int_x \left(\frac{-1}{\Box_x}\right)(\partial_\mu\, \mathrm{Tr}\log(\phi/\phi^\dagger))^2 . \tag{4.26}$$

It is not obvious how this expression should be read such that it does make sense. If it is equivalent to (4.25) then clearly no improvement has been achieved. The problem of a multivalued logarithm has merely been substituted by the problem of an infrared divergent integral in x space. Equation (4.26) is then a clear example of linguistical gymnastics that should be avoided: formally it appears to be chirally invariant, yet it is equivalent to the local term (4.25), which is not.

We conclude in this section that the aforementioned problem is not solved by the logarithmic potentials V'_a of eqs. (4.25), (4.26). Let us call this problem the "U(1) dilemma". The correct resolution of the U(1) dilemma will be given in the next sections.

5. Solution of the U(1) dilemma

We must keep in mind how and why an effective Lagrangian is constructed. The word "effective" is meant to imply that such a model is not intended to describe the system in all circumstances. Rather, the model gives a simplified treatment of the system in a given range of energies and momenta. In this case we are interested in energies and momenta lower than, say, 1 GeV.

Now the complete theory contains variables at much higher frequencies. In as far as they play a role at lower energies, we must assume that they have been taken care of in the effective model. Consequently, the simple identification (4.24) is not correct as it stands. It should be read as

$$\tilde{\phi} \simeq (\bar{q}_R q_L)_{\text{low frequencies}} . \tag{5.1}$$

But what does "low frequency" mean? In a gauge theory the concept "frequency" need not be gauge-invariant. Therefore the splitting between high frequency and low frequency components of the quark fields must depend in general on the gluonic fields A_μ. This is why the contribution of the high frequency components of the quark fields to the axial current $J_{\mu 5}$ may depend explicitly on the A fields, a fact that is correctly expressed by the so-called "anomalous commutators" of [6, 7]. After integrating out the high frequency modes of the quark fields, but before integrating out the A fields, we have an expression for the axial current which has the following form:

$$J_{\mu 5} = 2\mathrm{i}\, \mathrm{Tr}\{(\partial_\mu \phi^*)\, \phi - \phi^* \partial_\mu \phi\} + J'_{\mu 5}(A) . \tag{5.2}$$

It is $J'_{\mu 5}(A)$ which is responsible for the nontrivial axial charge of the quantity $F\tilde{F}$ in (4.23). Let Q'_5 be the charge corresponding to $J'_{\mu 5}$. How does $F\tilde{F}$ commute with Q'_5?

Rather than $F\tilde{F}$ itself, it is the integral over some space-time volume ΔV,

$$\int_{\Delta V} F\tilde{F} = \nu^{+}_{\Delta V} - \nu^{-}_{\Delta V}\,, \tag{5.3}$$

that is relevant in (4.13). (We use the short hand notation of eq. (3.4).) Let us take ΔV so small that

$$\int_{\Delta V} F\tilde{F} = 0 \quad \text{or} \quad \pm 1\,. \tag{5.4}$$

(i) If $F\tilde{F} = 0$ then we are not interested in its quantum numbers.
(ii) Whenever the right hand side of (5.3) is ±1 we have an amplitude in which $\pm 2L$ units of axial charge are created.
(iii) The higher values of the right hand side are negligible.

The creation or annihilation of axial charges occurs because of the extra *high frequency* modes of $\bar{\psi}_L$, ψ_R or ψ_L and $\bar{\psi}_R$ that make the functional integral non-invariant. Let us call their contribution Z. If $\nu^{+}_{\Delta V} = 1$ then

$$Z = \int \prod_N \mathrm{D}\psi_L \prod_{N+1} \mathrm{D}\psi_R \prod_N \mathrm{D}\bar{\psi}_R \prod_{N+1} \mathrm{D}\bar{\psi}_L (\exp S)\,, \tag{5.5}$$

where the subscripts under the multiplication symbols denote the numbers of variables to be integrated over. Then if

$$\begin{aligned} \psi_L &\to U_L \psi_L = \mathrm{e}^{-\mathrm{i}\omega} \psi_L\,, \\ \psi_R &\to U_R \psi_R = \mathrm{e}^{\mathrm{i}\omega} \psi_R\,, \end{aligned} \tag{5.6}$$

we have for all integrals over the anticommuting fields:

$$\begin{aligned} \int \mathrm{D}\psi_L &\to \mathrm{e}^{\mathrm{i}\omega} \int \mathrm{D}\psi_L\,, \\ \int \mathrm{D}\psi_R &\to \mathrm{e}^{-\mathrm{i}\omega} \int \mathrm{D}\psi_R\,, \end{aligned} \tag{5.7}$$

so that

$$Z \to \mathrm{e}^{-2\mathrm{i}\omega L} Z\,. \tag{5.8}$$

In this discussion we only include the high frequency components of the ψ fields. We see two things: the effective interaction Z due to an instanton transforms exactly as our insertion U_a of eq. (2.13), and secondly that, in a simplified picture where $F\tilde{F}$ takes integer values only (eq. (5.3)), the quantity $F\tilde{F}$, *after integration over the high frequency fermionic modes*, transforms with a factor

$$\mathrm{e}^{\pm 2\mathrm{i}\omega L}\,, \tag{5.9}$$

so that there is no longer any conflict* with (4.23). The transformation rules (4.1–4.8) hold for the effective fields. The terms containing Λ_5^0 tell us how the various fields commute with Q_5:

$$[Q_5, \Psi_L] = \Psi_L \tag{5.10}$$

$$[Q_5, \psi_R] = -\psi_R \tag{5.11}$$

$$[Q_5, \phi] = 2\phi \tag{5.12}$$

$$[Q_5, \sigma] = 2i\eta \tag{5.13}$$

etc.

6. Fictitious symmetry

The chiral U(1) symmetry breaking in QCD is an explicit one because the functional measure $\Pi\, D\psi$ fails to be chirally invariant when regularized in a gauge-invariant way [21]. This neatly explains why no massless Goldstone bosons are associated with this symmetry. Yet in several treatizes the words "spontaneous symmetry breaking" are used. How can this be?

Any broken global symmetry can formally be considered as a "spontaneously" broken one by a procedure consisting of two steps.

(i) Enlarge the physically accessible Hilbert space by adding all those Hilbert spaces of systems that would be obtained by applying the phoney symmetry transformation:

$$\mathfrak{H}' = \mathfrak{H} \times S \tag{6.1}$$

where $\mathfrak{H}$ is the original Hilbert space and S the space of physical constants describing symmetry breaking.

(ii) Define the symmetry operator(s) as acting both in S and in $\mathfrak{H}$. We then obtain transformations in $\mathfrak{H}'$ that obviously leave the Hamiltonian H invariant. This procedure allows one to write down Ward identities for theories with symmetries broken explicitly by one or more terms in the Lagrangian. Since such identities were excessively used and advocated by Veltman in his early work on gauge theories with mass-insertions, we propose to refer to the above transformations as Veltman transformations [12].

Consider for example quantum electrodynamics with electron mass term

$$-m^*\bar{\psi}_L\psi_R - m\bar{\psi}_R\psi_L\,. \tag{6.2}$$

Then S is the space of complex numbers m. In this theory then, m is promoted to be an operator rather than a c-number. The chiral transformation

$$\begin{aligned} \psi_{\substack{R\\L}} &\to e^{\pm i\omega}\psi_{\substack{R\\L}} \\ m &\to e^{2i\omega} m \end{aligned} \tag{6.3}$$

* For the factor $e^{i\theta}$ see section 7.

is obviously an invariance of this theory. If in the "physical world"

$$\langle m \rangle = m = \text{real} \tag{6.4}$$

then one could argue that the symmetry (6.3) is "spontaneously broken".

The canonical charge operator $\tilde{Q}$ associated with (6.3) is now

$$\tilde{Q} = Q_R - Q_L + 2\left(m^* \frac{\partial}{\partial m^*} - m \frac{\partial}{\partial m}\right), \tag{6.5}$$

which commutes with (6.2). Thus, $\tilde{Q}$ is exactly conserved. But, since m is not a dynamical field, the new term cannot be written as an integral over 3-space, unless we enlarge the Hilbert space once again.

Let us now consider a Feynman diagram in which the mass term (6.2) occurs perturbatively as a two-prong vertex. Let there be a diagram with ν^+ insertions of the last term in (6.2), going with m, and ν^- of the first term, going with m^*. We have

$$\begin{aligned} &\Delta Q_5 = \Delta Q_R - \Delta Q_L = 2(\nu^+ - \nu^-) \\ &\Delta \tilde{Q} = 0 . \end{aligned} \tag{6.6}$$

Only by brute force one could produce a current of which the fourth component would give a charge satisfying (6.6):

$$\begin{aligned} &\tilde{J}_\mu = J_{\mu 5} + K_\mu \\ &\partial_\mu K_\mu(x) = -2\mathrm{i}(\rho^+(x) - \rho^-(x)) \end{aligned} \tag{6.7}$$

where $\rho^{\pm}(x)$ is the density of the corresponding mass insertion vertices,

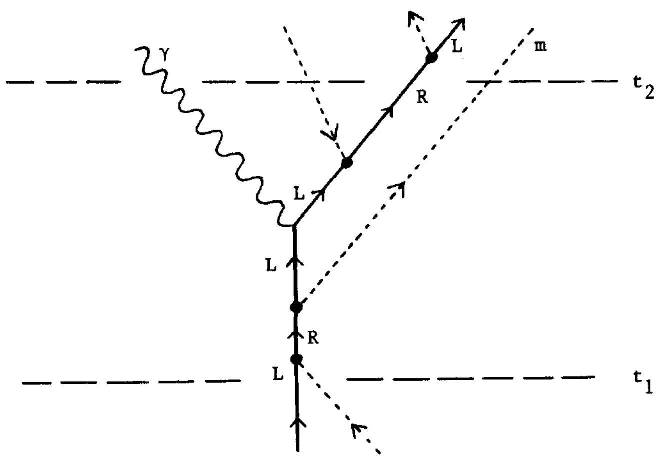

Fig. 1. Propagating electron (solid line) with mass terms. The propagators are expanded in m, yielding artificial particles (schizons, dotted lines) that carry away two units of axial charge, but no energy-momentum. Total chiral charge Q_5 is conserved. Here $Q_5(t_1) = Q_5(t_2) = 1$.

$$K_\mu(x) = -2\mathrm{i}\Box^{-1}\partial_\mu(\rho^+ - \rho^-)\,. \tag{6.8}$$

Clearly, K_μ is not locally observable. There is a nonobservable "Goldstone ghost" (the pole of $\Box^{-1}$). It goes without saying that $\tilde{Q}$, although exactly conserved, and $\tilde{J}_\mu$, are not very useful for canonical formalism. Yet the current $J^L_{\mu 5,\mathrm{sym}}$ and the charge Q^L_5 as used in ref. [7] are precisely of this form. This will be explained in the next section.

A neat way to implement the symmetry (6.3) is to treat the parameter m as a field: the "schizon", or "spurion", as those auxiliary objects are sometimes called to describe explicit symmetry breaking, such as isospin breaking by electromagnetism. The schizon field has a nonvanishing vacuum expectation value (6.4). Diagrammatically, a propagating electron could be represented by a diagram (fig. 1). Defining $\tilde{Q} = \pm 2$ for the schizons we see that $\tilde{Q}$ is absolutely conserved. Of course $\tilde{Q}$ is also "spontaneously broken".

7. The "exactly conserved chiral charge" in a canonically quantized theory

The fictitious symmetry described in the previous section can be mimicked in a gauge theory in a way that looks very real. Consider instead of (4.11), the current

$$\begin{aligned} J_{\mu 5,\mathrm{sym}} &= J_{\mu 5} + K_\mu \\ K_\mu &= -\frac{g^2 \mathrm{i} L}{16\pi^2}\,\varepsilon_{\mu\nu\alpha\beta} A^a_\nu(\partial_\alpha A^a_\beta + \tfrac{1}{3} g f_{abc} A^b_\alpha A^c_\beta)\,. \end{aligned} \tag{7.1}$$

Then, in the limit $m \to 0$, one has

$$\partial_\mu J_{\mu 5,\mathrm{sym}} = 0\,. \tag{7.2}$$

The corresponding charge,

$$Q_{5,\mathrm{sym}} = \int J_{05,\mathrm{sym}}\,\mathrm{d}^3\boldsymbol{x} \tag{7.3}$$

generates "exact" chiral transformations. How does this operator act in Hilbert space?

To answer this question we must formulate the canonical quantization of the gluon field carefully. Conceptually the most transparent way is to first choose the temporal gauge:

$$A_0 = 0\,, \tag{7.4}$$

which leaves us formally the set of all states $|A(\boldsymbol{x}), \psi(\boldsymbol{x}), \bar{\psi}(\boldsymbol{x})\rangle$ at a given time t, where ψ and $\bar{\psi}$ should be seen as Grassmann numbers. Let us call the Hilbert space spanned by all these states the "huge" Hilbert space.

Then (7.4) leaves us invariance under all time-independent gauge transformations $\Omega = \Omega(\boldsymbol{x})$, so that the Hamiltonian in this space is invariant under a group G composed of gauge transformations $\Omega(\boldsymbol{x})$ that may vary from point to point. This generates an invariance at each $\boldsymbol{x}$, according to Noether's theorem. Writing

$$\Omega|A,\psi,\bar\psi\rangle=|A^{\Omega}\psi^{\Omega}\bar\psi^{\Omega}\rangle \tag{7.5}$$

where the subscript Ω indicates how the fields are gauge-transformed, we have

$$[H,\Omega]=0\,. \tag{7.6}$$

We can impose the gauge conditions of the second type:

$$\Omega|\Psi\rangle=|\Psi\rangle\,, \tag{7.7}$$

for all *infinitesimal* Ω, acting nontrivially only in a finite region of 3-space:

$$A^{\Omega}_{\mu}(\boldsymbol{x})=A(\boldsymbol{x})+\mathrm{D}_{\mu}\Lambda(\boldsymbol{x})\,,$$

Λ infinitesimal, and with compact support.

States $|\Psi\rangle$ satisfying (7.7) are said to be in the "large" Hilbert space (which is not as large as the "huge" one).

Finally, we consider all Ω with nontrivial winding number ν

$$\Omega_{\nu}|\Psi\rangle=\mathrm{e}^{\mathrm{i}\theta\nu}|\Psi\rangle\,. \tag{7.8}$$

These states $|\Psi\rangle$ are said to constitute the small, or physical Hilbert space at given θ.

Now notice that $J_{\mu5,\mathrm{sym}}$ does not commute with Ω:

$$[J_{\mu5,\mathrm{sym}},\Omega]=\frac{\mathrm{i}Lg^{2}}{16\pi}\,\mathrm{D}_{\nu}\Lambda\cdot\varepsilon_{\mu\nu\alpha\beta}F_{\alpha\beta}\,. \tag{7.9}$$

Therefore, $J_{\mu5,\mathrm{sym}}$ cannot be considered to be an operator for states in the "large" Hilbert space. Acting on a state satisfying (7.7) it produces a state not satisfying (7.7).

Now the charge $Q_{5,\mathrm{sym}}$ of eq. (7.3) *does* commute with all Ω with $\nu=0$, but not with the others:

$$[Q_{5,\mathrm{sym}},\Omega_{\nu}]=2\mathrm{i}L\nu\Omega_{\nu}\,. \tag{7.10}$$

Therefore, $Q_{5,\mathrm{sym}}$ does act as an operator in the large Hilbert space, but not in the physical Hilbert space, because it mixes different θ values. We can write

$$\left[Q_{5,\mathrm{sym}}-2\mathrm{i}L\,\frac{\partial}{\partial\theta}\,,\,\Omega_{\nu}\right]=0\,. \tag{7.11}$$

We see that in every respect $Q_{5,\mathrm{sym}}$ behaves as $\tilde{Q}$ of the previous section, and $J_{\mu5,\mathrm{sym}}$ as $\tilde{J}$.

A Goldstone boson would emerge in the theory if, besides the states satisfying (7.8), it could be possible to construct physical states in which θ would depend on space-time:

$$\theta\overset{?}{=}\theta(\boldsymbol{x},t)\,. \tag{7.12}$$

Here it is obvious that (7.12) would be in contradiction with (7.7) and (7.8): If we would compare different Ω_ν but with the same ν, such that the support of $\Omega^{(1)}$ would be in a region near $x^{(1)}$, and that of $\Omega^{(2)}$ near $x^{(2)}$, then the combination

$$\Omega^{(1)}\Omega^{(2)-1}$$

would have winding number zero. So the second gauge constraint would exclude any states for which $\theta(x^{(1)})\neq\theta(x^{(2)})$. This is an important contrast with systems such as a ferromagnet, where local fluctuations are allowed, which, because of their large correlation lengths, correspond to massless excitations.

Because of the similarity between (7.11) and (6.5) we can consider $e^{i\theta}$ as a "schizon" field just as the electron mass term. Since θ cannot have any space-time dependence this schizon field cannot carry away any energy or momentum, just as m in the previous section.

Although $Q_{5,\mathrm{sym}}$ does not act in the "physical" Hilbert space, it is possible to write Ward identities [6, 7] due to its formal conservation,

$$\int \mathrm{d}^4x\, \partial_\mu T\langle J^L_{\mu 5}(x), \mathrm{Op}\rangle = 2L\int \mathrm{d}^4x\, \partial_\mu T\langle K_\mu(x), \mathrm{Op}\rangle + \int \mathrm{d}^4x\, T\langle D_L(x), \mathrm{Op}\rangle + \langle[Q_{5,\mathrm{sym}}, \mathrm{Op}]\rangle\,, \tag{7.13}$$

where Op stands for any operator; and

$$D_L = 2\mathrm{i}m\bar{\psi}\gamma_5\psi\,, \tag{7.14}$$

which will vanish when $m\to 0$. K_μ satisfies

$$\partial_\mu K_\mu = -\frac{g^2\mathrm{i}}{32\pi^2}\,F^a_{\mu\nu}\tilde{F}^a_{\mu\nu}\,. \tag{7.15}$$

One can take

$$\mathrm{Op} = K_\nu(0)\,, \tag{7.16}$$

and assume

$$[Q_{5,\mathrm{sym}}, K_\nu] = 0 \tag{7.17}$$

while putting $D_L\to 0$. (7.18)

Now (7.17) is not obvious. Substituting (7.15) gives

$$[Q_{5,\mathrm{sym}}, F\tilde{F}] = 0\,. \tag{7.19}$$

On the other hand we showed in section 5 that $F\tilde{F}$ has nontrivial chiral transformation properties. This however corresponds to

$$[Q_5, F\tilde{F}]\neq 0\,. \tag{7.20}$$

Indeed,

$$\langle [Q_5, F\tilde{F}] \rangle \cong C \cdot f^2 \tag{7.21}$$

where the right hand side follows from the substitution (4.23) and the commutation rule (5.12). C is a constant and f the σ expectation value.

Equation (7.19) is a fundamental starting point of the discussions in refs. [6–8]. The difference between (7.19) and (7.20) must apparently be made up by the contribution of K_0 to $Q_{5,\mathrm{sym}}$. Now K_0 is not a physically observable field. Assigning to it the conventional commutation rules to be deduced from its composition in terms of color gauge fields is only allowed if one works in the "huge" Hilbert space including the gauge noninvariant states.

All we have to do to incorporate the fictitious symmetry generated by $Q_{5,\mathrm{sym}}$ into our model of effective fields described in section 2, is to add a schizon field, enlarging the Hilbert space. Let us call the schizon field

$$S = \mathrm{e}^{\mathrm{i}\theta} . \tag{7.22}$$

Our new identification is

$$F\tilde{F} = \frac{128\pi^2 \mathrm{i}}{Lg^2} \operatorname{Im}(S \det \phi) , \tag{7.23}$$

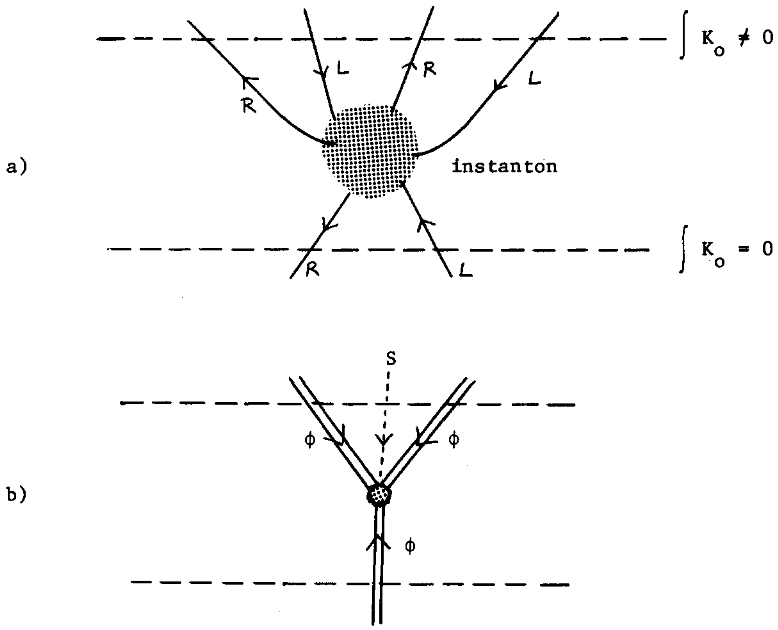

Fig. 2. Effective instanton action and its Q_5 symmetry properties. (a) Due to fermionic zero modes $2L$ units of Q_5 are absorbed at the site of the instanton. In the same time the charge generated by K_0 is not conserved. (b) In the effective theory the fermions are replaced by the ϕ field, and K_0 by a schizon S. The ϕ carry 2 units of Q_5 each, and S has $-2L$ units. S has a nonvanishing vacuum expectation value.

and if we postulate, in addition to (5.10)–(5.13) for $Q_{5,\mathrm{sym}}$ also (see 7.11):

$$[Q_{5,\mathrm{sym}}, S] = -2LS\,, \tag{7.24}$$

then with (5.12) we find that $F\tilde{F}$ commutes with $Q_{5,\mathrm{sym}}$. Substituting $e^{i\theta}\det\phi$ by $S\det\phi$ in (2.13) we see that indeed our effective field theory obeys the fictitious symmetry generated by $Q_{5,\mathrm{sym}}$. It must therefore also obey the so-called anomalous Ward identities. See fig. 2.

8. Diagrammatic interpretation of the anomalous Ward identities

The conclusion of the previous sections is that the model of section 2, with the substitution

$$e^{i\theta} \rightarrow S \tag{8.1}$$

obeys all anomalous Ward identities. It also exhibits in a very transparent way how the symmetries are now spontaneously broken. There are two vacuum expectation values:

$$\langle S\rangle = e^{i\theta} \tag{8.2}$$

$$\langle\sigma\rangle = f\,. \tag{8.3}$$

Both break $Q_{5,\mathrm{sym}}$ conservation. We can now draw Feynman diagrams in the Wigner representation by explicitly adding the vacuum bubbles due to (8.2) and (8.3). See fig. 3. By summing over the bubble insertions (geometric series which are trivial to sum), one reobtains the Goldstone representation of the particles. The σ bubbles tend to make the pions and eta massless, but the terms with κ (and the quark masses m) contribute linearly to m^2 for the various mesons. These diagrams clearly visualize where the masses come from and how the $Q_{5,\mathrm{sym}}$ charges are absorbed into the vacuum.

In ref. [7] an apparent problem was raised by their equations (4.27) and (4.28): they suggest the

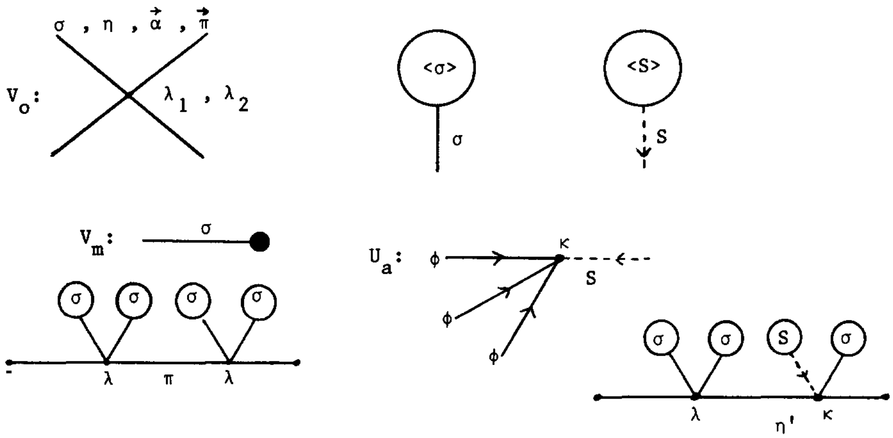

Fig. 3. Feynman rules in the Wigner mode. The σ blobs render the pion and the η massless. But the S blob gives a mass to η. Here we drew explicitly the η' propagator ($L=3$). Its mass comes from the insertion at the right. Note that $m^2_{\eta'}$ is linear in $\langle S\rangle$ and in $\langle\sigma\rangle$: $m^2_{\eta'} \propto \kappa f$.

need for field configurations with fractional winding number ν, which would correspond with the breaking up of our S field into components with smaller $Q_{5,\mathrm{sym}}$ charges. In our diagrams we clearly see that there is no such need. If we have a Green function with an operator that creates only two chiral charges,

$$\chi_{\mathrm{Op}} = 2$$

then the sigma field can absorb these two, or add $2(L-1)$ more and have them absorbed by S. The vacuum simply isn't an eigenstate of Q_5 (nor $Q_{5,\mathrm{sym}}$) as it was assumed.

9. Conclusion

The disagreements between the approach of the Crewther's school to the U(1) problem, using anomalous Ward identities, and the more standard beliefs are not as wide as they appear. Their anomalous Ward identities, if applied with appropriate care, are perfectly valid for a simple effective field theory that clearly exhibits the most likely vacuum structure of QCD. It is important however to realize that the relevant exactly conserved chiral charge $Q_{5,\mathrm{sym}}$ is not physically observable, something which explains the need for the introduction of a spurion field S in the effective theory. It appears that the consequences of working with an unphysical symmetry were underestimated in ref. [7]. Some of the difficulties signalled in [7] were due to the too strong assumption that the vacuum is an eigenstate of Q_5. That such assumptions are unnecessary and probably wrong would have been realized if they had taken the effective theory more seriously.

The fact that the effective theory of section 2 displays the correct symmetry properties does not have to mean that it is accurate. Indeed it could be that instantons tend to split into "merons" [16], a dynamical property that might be a factor in the spontaneous chiral symmetry breaking mechanism [14]. But these aspects do not affect the symmetry transformation properties of the fields under consideration. More fields, describing higher resonances, could have to be added. The baryonic degrees of freedom are most likely to be considered as extended solutions of the effective field equations (skyrmions). That these skyrmions [17] indeed possess the relevant baryonic quantum numbers was discovered by Witten [18].

The author thanks R.J. Crewther for his patience in extensive discussions, even though no complete agreement was reached.

Appendix A. The sign of ΔQ_5

In ref. [7] the present author's work was claimed to be in error at various places. Although some minor technical corrections on the computed coefficients in the quantum corrections due to instantons were found (see ref. [20] and appendix B) and even some insignificant inaccuracies in the notation of a sign might occur, we stress here that none of those claims of ref. [7] were justified. In particular there are no fundamental discrepancies in the sign of ΔQ_5.

Let us here ignore the masses of the quarks. As formulated in section 4, an instanton in a finite space-time volume V causes a transition with

$$\Delta Q_5 = 2L \tag{A1}$$

where Q_5 is the gauge-invariant axial charge. Since $Q_{5,\mathrm{sym}}$, as defined in section 7, is now strictly conserved one obviously has

$$\Delta Q_{5,\mathrm{sym}} = 0\,, \tag{A2}$$

which is of course only defined in the "large" Hilbert space comprising all θ worlds.

Instead of (A1–A2), we read in ref. [7]:

$$\Delta Q_5 = 0\,; \qquad \Delta Q_{5,\mathrm{sym}} = -2L\,. \tag{A3}$$

These are not the properties of a closed space-time volume such as we described, but represent the features of a Green's function where the asymptotic states are θ-vacua. Since $Q_{5,\mathrm{sym}}$ contains explicitly the operator $\partial/\partial\theta$ in the "large" Hilbert space (cf. eq. (7.11)), the θ-vacuum is not invariant under $Q_{5,\mathrm{sym}}$. This is why (A3) is not in conflict with $Q_{5,\mathrm{sym}}$ conservation. But we also see that (A1–A2) and (A3) hold under different boundary conditions (the reason why $\Delta Q_5 = 0$ for Green's functions in a θ-vacuum is correctly explained in ref. [7]).

Appendix B. The amplitude of the instanton effects

To get even a rough estimate of the size of our instanton's contribution to an amplitude requires lengthy calculations. A detailed account would require a complete reprinting of this author's work in ref. [19]. As pointed out in [20] there were some minor errors in the first publication which we will discuss briefly here.

Let $A_\mu^{a\,\mathrm{inst}}$ be the field of one instanton. Then the amplitude due to one instanton in Euclidean space in a short hand notation is:

$$W = {}_{\mathrm{out}}\langle\,|\,\rangle_{\mathrm{in}} = \int \mathrm{D}A^{\mathrm{qu}}\,\mathrm{D}\psi \cdots \exp \int \mathrm{d}^4x\, \mathfrak{L}(A^{\mathrm{inst}} + A^{\mathrm{qu}}, \psi, \bar{\psi})\,. \tag{B1}$$

We have

$$\int \mathrm{d}^4x\, \mathfrak{L}(A^{\mathrm{inst}}) = -8\pi^2/g^2 \tag{B2}$$

and the part of the action that is quadratic in A^{qu}, ψ, $\bar{\psi}$, is

$$\mathfrak{L}^{(2)} = -\tfrac{1}{2}(\mathrm{D}_\mu A_\nu^{\mathrm{qu}})^2 + \tfrac{1}{2}(\mathrm{D}_\mu A_\mu^{\mathrm{qu}})^2 - g A_\nu^{a\,\mathrm{qu}} f_{\mathrm{abc}} G_{\mu\nu}^{b\,\mathrm{inst}} A_\mu^{c\,\mathrm{qu}} - \bar{\psi}\gamma_\mu \mathrm{D}_\mu \psi - \tfrac{1}{2} C_1^2 + \mathfrak{L}_1^{\mathrm{ghost}} + \bar{\psi} J \psi\,, \tag{B3}$$

where D_μ is the covariant derivative with respect to A^{inst} and $-\frac{1}{2}C_1^2$ is the gauge fixing term, producing a ghost described by $\mathfrak{L}_1^{\mathrm{ghost}}$. We added a source term for the fermions.

Now the integral (B1) in lowest nontrivial order is Gaussian with nontrivial coefficients. Formally we can write the outcome as

$$W = \exp(-8\pi^2/g^2)(\det \mathfrak{M}_A)^{-1/2} \det \mathfrak{M}_\psi \det \mathfrak{M}_{\rm gh} \tag{B4}$$

and the determinants can be computed by diagonalization:

$$\mathfrak{M}_i \psi = E_i \psi . \tag{B5}$$

It is clear that (B4) is highly divergent unless we formulate very precisely an appropriate subtraction procedure. A convenient method is to apply first a variety of Pauli–Villars regularization to the fields $A^{\rm qu}$ and ψ and then add correction terms to be obtained by comparing this regularization scheme to for instance dimensional regularization.

It is then found that (B4) is not the complete answer: there are zero eigenvalues of $\mathfrak{M}_A$, $\mathfrak{M}_\psi$ and $\mathfrak{M}_{\rm gh}$ which have to be considered separately. The corresponding eigenmodes must be replaced by collective coordinates including an appropriate Jacobian for this transformation. Since (B4) must be compared with the vacuum transition (in absence of instantons) the collective coordinate integration is to be divided by a norm factor determined by a Gaussian integral. In our work [19] this factor was taken to be $\sqrt{\pi}$. However, the relevant Gaussian integral was

$$\int \exp(-\tfrac{1}{2}x^2)\,{\rm d}x = \sqrt{2\pi} , \tag{B6}$$

and the factor 2 was missed. Consequently, in the final expressions (12.1), (12.5), (13.8) and (15.1) we must replace 2^{14} by 2^{10}.

In comparing the Pauli–Villars regulators with dimensional regularization, section 13 of ref. [19], another error was made. Applying the two regulators to the integral

$$\frac{1}{(2\pi)^n} \int {\rm d}^n k \, \frac{1}{(k^2 + \mu_0^2)^2} \tag{B7}$$

we find

$$\frac{1}{(2\pi)^2} \left(\frac{2}{4-n} - \gamma - 2\log \mu_0 + \log 4\pi + \mathcal{O}(4-n) + \mathcal{O}\left(\frac{1}{\mu_0^2}\right)\right) , \tag{B8}$$

so, if μ_0 is a regulator mass, then the comparison yields

$$\log \mu_0 \to \frac{1}{4-n} - \tfrac{1}{2}\gamma + \tfrac{1}{2}\log 4\pi . \tag{B9}$$

The integral (13.5) of ref. [19] should be replaced by

$$\frac{1}{(2\pi)^n} \int {\rm d}^n k \, \frac{k_\mu k_\nu k_\alpha k_\beta}{(k^2 + \mu_0^2)^2} , \tag{B10}$$

which gives

$$\frac{1}{(4\pi)^2 \cdot 4!} \left[\frac{2}{4-n} - \gamma - 2\log \mu_0 + \log 4\pi + \mathcal{O}(4-n) + \mathcal{O}\left(\frac{1}{\mu_0^2}\right)\right] (\delta_{\mu\nu}\delta_{\alpha\beta} + \delta_{\mu\alpha}\delta_{\nu\beta} + \delta_{\mu\beta}\delta_{\nu\alpha}) . \tag{B11}$$

We see the same substitution (B9) can be used and there is no correction term 5/6 as written in [14]. Equation (B9) relates the two regulators in all circumstances.

On the other hand the number of A_μ^{qu} fields in dimensional regularization is n rather than 4. So for these fields we need the substitution

$$\log \mu_0 \to \frac{1}{4-n} - \tfrac{1}{2}\gamma + \tfrac{1}{2}\log 4\pi - 1\,. \tag{B12}$$

The rest of the procedure to sum all contributions of the eigenmodes of the operators $\mathfrak{M}_A$, $\mathfrak{M}_\psi$, $\mathfrak{M}_{\mathrm{gh}}$ and possible scalar contributions $\mathfrak{M}_s$, and the contributions of the collective coordinates is all as explained in [19]. The result is now the effective Lagrangian (for the case that the color gauge group is SU(2)):

$$\begin{aligned}\mathfrak{L}^{\mathrm{eff}}(z) = 2^{10+3N^{\mathrm{f}}}\pi^{6+2N^{\mathrm{f}}}g^{-8}\int \rho^{3N^{\mathrm{f}}-5}\,\mathrm{d}\rho\,\exp\Big\{&-\frac{8\pi^2}{[g_{\mathrm{R}}^{\mathrm{D}}(\mu)]^2}\\ &+\log(\mu\rho)\Big[\frac{22}{3}-\frac{1}{6}\sum_t N^{\mathrm{s}}(t)C(t)-\tfrac{2}{3}N^{\mathrm{f}}\Big]+A-\sum_t N^{\mathrm{s}}(t)A(t)-N^{\mathrm{f}}B\Big\}\\ &\times\Big\langle\prod_{s=1}^{N^{\mathrm{f}}}(\bar\psi_s\omega)(\bar\omega\psi_5)\Big\rangle+\mathrm{h.c.}\,,\end{aligned} \tag{B13}$$

where N^{f} is the number of fermions in the doublet representation, ρ is a scalar parameter for the instanton, μ is an arbitrary mass unit enabling us to obtain a renormalization group invariant expression, and $N^{\mathrm{s}}(t)$ is the number of scalar field representations with color t.

Defining coefficients $\alpha(t)$ as in table 1, we have now:

$$\begin{aligned} A &= -\alpha(1) + \tfrac{11}{3}(\log 4\pi - \gamma) + \tfrac{1}{3} = 7.053\,991\,03\\ A(t) &= -\alpha(t) + \tfrac{1}{12}(\log 4\pi - \gamma)C(t)\\ A(1/2) &= 0.308\,690\,69\\ A(1) &= 1.094\,576\,62\\ A(3/2) &= 2.481\,356\,10\\ B &= -2\alpha(1/2) + \tfrac{1}{3}(\log 4\pi - \gamma) = 0.359\,522\,90\,. \end{aligned} \tag{B14}$$

Table 1

t	$C(t)$	$\alpha(t)$
0	0	0
1/2	1	$2R - \frac{1}{6}\log 2 - 17/72$
1	4	$8R + \frac{1}{3}\log 2 - 16/9$
3/2	10	$20R + 4\log 3 - \frac{5}{2}\log 2 - 265/36$

$R = \frac{1}{12}(\log 2\pi + \gamma) + \frac{1}{2\pi^2}\sum_2^\infty \frac{\log s}{s^2} = 0.248\,754\,477.$

Finally, the spinors ω are normalized by

$$\sum_{\alpha} \omega_\alpha \bar{\omega}_\alpha = \tfrac{1}{2}(1+\gamma_5) \tag{B15}$$

and required to be smeared in color space, such that for instance

$$\langle \omega_\alpha \bar{\omega}_\beta \rangle = \tfrac{1}{4}\delta_{\alpha\beta}(1+\gamma_5) \tag{B16}$$

and, in the case $L = N^{\mathrm{f}} = 2$:

$$\left\langle \prod_{s=1}^{2} (\bar{\psi}_s \omega)(\bar{\omega}\psi_s) \right\rangle = \frac{1}{24}\,(2\delta^{\beta_1}_{\alpha_1}\delta^{\beta_2}_{\alpha_2} - \delta^{\beta_2}_{\alpha_1}\delta^{\beta_1}_{\alpha_2})\varepsilon^{st}\bar{\psi}^{\alpha_1}_1(1+\gamma_5)\psi^{\beta_1}_s\bar{\psi}^{\alpha_2}_2(1+\gamma_5)\psi^{\beta_2}_t \tag{B17}$$

where s and t are flavor indices and α_i, β_i color indices.

The ρ integral may seem to diverge in most interesting cases ($N^{\mathrm{f}} > 1$). Note however that it would be natural to choose

$$\mu = 1/\rho \tag{B18}$$

and substitute g^{-8} by the running value $g(\mu)^{-8}$. At large ρ one might take $g \propto \rho$ and thus improve the convergence. Of course the infrared end of the integral is quite uncertain because in our perturbative procedure the effects of confinement etc. have not been taken into account. This inhibits a precise evaluation of the amplitude. A rough estimate (for a color SU(2) theory) is obtained if we take at large ρ

$$g^2(1/\rho) \to 16\pi\rho^2\sigma \tag{B19}$$

where σ is the string constant. Then quarks with color charge 1/2 at a distance ρ from each other feel a force

$$\sigma = \tfrac{1}{4} g^2/4\pi\rho^2\,. \tag{B20}$$

Our integral becomes, in the case $N^{\mathrm{f}} = 2$,

$$\mathfrak{L}^{\mathrm{eff}} = 2^{16}\pi^{10}\,\mathrm{e}^{A-2B}\mathfrak{L}_1 \int \frac{\rho\,\mathrm{d}\rho}{g^8(1/\rho)} \exp - 8\pi^2/g^2(1/\rho)\,. \tag{B21}$$

From (B19) we get

$$\rho\,\mathrm{d}\rho \cong g\,\mathrm{d}g/16\pi\sigma \tag{B22}$$

and using $x = 1/g^2$ the integral in (B21) is

$$\frac{1}{32\pi\sigma}\int_0^\infty \mathrm{d}x\, x^2\, \mathrm{e}^{-8\pi^2 x} = 2^{-13}\pi^{-7}\sigma^{-1} \tag{B23}$$

so that

$$\mathfrak{L}^{\text{eff}} = 8\pi^3 \, e^{A-2B}\sigma^{-1}\mathfrak{L}_1 \tag{B24}$$

where $\mathfrak{L}_1$ is the Lagrangian (B17).

This result is uncomfortably large, but then the approximations used here (eq. B19) could at best only be expected to yield the order of magnitude of the expected interaction, which is clearly a strong one. Note that we used a minimal subtraction scheme that included the term $\log 4\pi$ in (B9). If we left it out then (B24) would be reduced by a factor $(4\pi)^{-3} = 2^{-6}\pi^{-3}$. This is just to illustrate how sensitively the amplitude obtained depends upon the assumptions.

References

[1] S.L. Glashow, in: Hadrons and their interactions, Erice 1967, ed. A. Zichichi (Acad. Press, New York, 1968) p. 83;
S.L. Glashow, R. Jackiw and S.-S. Shei, Phys. Rev. 187 (1969) 1916;
M. Gell-Mann, in: Proc. Third Topical Conf. on Particle Physics, Honolulu 1969, eds. W.A. Simonds and S.F. Tuan (Western Periodicals, Los Angeles, 1970) p. 1, and in: Elementary Particle Physics, Schladming 1972, ed. P. Urban (Springer-Verlag, 1972); Acta Physica Austriaca Suppl. IX (1972) 733;
H. Fritzsch and M. Gell-Mann, Proc. XVI Intern. Conf. on H.E.P., Chicago 1972, eds. J.D. Jackson and A. Roberts, Vol. 2, p. 135, and Phys. Lett. 47B (1973) 365.
[2] S.L. Adler, Phys. Rev. 177 (1969) 2426;
J.S. Bell and R. Jackiw, Nuovo Cim. 60A (1969) 47;
S.L. Adler and W.A. Bardeen, Phys. Rev. 182 (1969) 1517.
[3] A.A. Belavin, A.M. Polyakov, A.S. Schwartz and Yu.S. Tyupkin, Phys. Lett. 59B (1975) 85.
[4] G. 't Hooft, Phys. Rev. Lett. 37 (1976) 8; Phys. Rev. D14 (1976) 3432;
R. Jackiw and C. Rebbi, Phys. Rev. Lett. 37 (1976) 172;
C.G. Callan Jr., R.F. Dashen and D.J. Gross, Phys. Lett. 63B (1976) 334; Phys. Rev. D17 (1978) 2717.
[5] S. Coleman, in: The Whys of Subnuclear Physics, Erice 1977, ed. A. Zichichi (Plenum Press, New York, 1979) p. 805.
[6] R. Crewther, Phys. Lett. 70B (1977) 349; Riv. Nuovo Cim. 2 (1979) 63;
R. Crewther, in: Facts and Prospects of Gauge Theories, Schladming 1978, ed. P. Urban (Springer-Verlag, 1978); Acta Phys. Austriaca Suppl. XIX (1978) 47.
[7] G.A. Christos, Phys. Reports 116 (1984) 251.
[8] R. Crewther, private communication.
[9] J.C. Ward, Phys. Rev. 78 (1950) 1824;
Y. Takahashi, Nuovo Cim. 6 (1957) 370.
[10] A. Slavnov, Theor. and Math. Phys. 10 (1972) 153 (in Russian), transl. Theor. and Math. Phys. 10, p. 99;
J.C. Taylor, Nucl. Phys. B33 (1971) 436.
[11] C. Becchi, A. Rouet and R. Stora, Comm. Math. Phys. 42 (1975) 127; Ann. Phys. (N.Y.) 98 (1976) 287.
[12] M. Veltman, Nucl. Phys. B7 (1968) 637; Nucl. Phys. B21 (1970) 288.
[13] There is no paper by Bell or Treiman on this particular subject. See ref. [12].
[14] A.R. Zhitnitsky, The discrete chiral symmetry breaking in QCD as a manifestation of the Rubakov–Callan effect, Novosibirsk preprint 1986.
[15] C. Lee and W. Bardeen, Nucl. Phys. B153 (1979) 210.
[16] C.G. Callan Jr., R.F. Dashen and D.J. Gross, Phys. Lett. 66B (1977) 375; Phys. Rev. D17 (1978) 2717; Phys. Lett. 78B (1978) 307.
[17] T.H.R. Skyrme, Proc. Roy. Soc. A260 (1961) 127.
[18] E. Witten, Nucl. Phys. B223 (1983) 422; 433.
[19] G. 't Hooft, Phys. Rev. D14 (1976) 3432.
[20] F.R. Ore, Phys. Rev. D16 (1977) 2577;
G. 't Hooft, Phys. Rev. D18 (1978) 2199;
A. Hasenfratz and P. Hasenfratz, Nucl. Phys. B193 (1981) 210.
[21] K. Fujikawa, Phys. Rev. Lett. 42 (1979) 1195; Phys. Rev. D21 (1980) 2848.