THE STRUCTURE OF THE GAUGE THEORY VACUUM★

C.G. CALLAN, Jr.
Joseph Henry Laboratories, Princeton University, Prinecton, New Jersey 08540, USA

R.F. DASHEN*
Institute for Advanced Study, Princeton, New Jersey 08540, USA

and

D.J. GROSS
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540, USA

Received 20 May 1976

The finite action Euclidean solutions of gauge theories are shown to indicate the existence of tunneling between topologically distinct vacuum configurations. Diagonalization of the Hamiltonian then leads to a continuum of vacua. The construction and properties of these vacua are analyzed. In non-abelian theories of the strong interactions one finds spontaneous symmetry breaking of axial baryon number without the generation of a Goldstone boson, a mechanism for chiral SU(N) symmetry breaking and a possible source of T violation.

Polyakov [1] has recently pointed out that the Euclidean classical equations of motion of gauge theories have soliton-like solutions and has suggested that when properly included in the Euclidean functional integral they may have a bearing on the dynamics of confinement. The physical interpretation of these solutions has, however, been obscure since they are localized in time as well as space. In this letter we shall show that Euclidean gauge solitons describe events in which topologically distinct realizations of the gauge vacuum *tunnel* into one another and that this process radically changes the nature of the vacuum state. In fact, we find a continuum of vacua, each one of which is a superposition of the vacua with difinite topology and stable under the tunnelling process. The new vacua are the ground states of independent, and in general, inequivalent worlds (most striking, P and T are spontaneously violated in some of them!). When massless fermions are present, the vacuum tunnelling process forces a redefinition of the fermion vacuum as well and leads directly to spontaneous breakdown of chiral invariance without generating a "ninth" Goldstone boson. We have, in effect, shown that the vacuum "seizes" as suggested by Kogut and Susskind [2], and identified the mechanism by which it does so. Our primary aim in this letter will be to give arguments for the existence of the new vacuum structure and to present the correct form of the functional integral appropriate to studying the properties of a particular vacuum. In the spirit of displaying qualitative consequence of the new vacuum structure we shall also briefly summarize results obtained from rather crude approximations to the functional integral.

To explore the structure of the vacuum we study the Euclidean functional integral

$$\langle 0|\exp(-Ht)|0\rangle \underset{t\to\infty}{\longrightarrow} \int [DA_\mu D\psi] \times \exp\{-\int d^dx[\mathcal{L}(A_\mu, \psi_i \ldots) + \mathcal{L}_{gf}]\} \tag{1}$$

where d is the dimension of space time, $\mathcal{L}$ is the Langrange density of the theory, $\mathcal{L}_{gf}$ is a gauge-fixing term and the integration is to be done over all fields that approach vacuum values ($F_{\mu\nu} = 0$) at infinity. Now since $F_{\mu\nu} = 0$ implies $A_\mu = g^{-1}(x)\partial_\mu g(x)$ takes on values in the gauge group, G, any gauge field included in the functional integration defines a map of the sphere at Euclidean infinity into G. As pointed out by

★ Research supported by the National Science Foundation under Grant Number MPS 75-22514.
* Research sponsored in part by the ERDA under Grant No. E(11-1)-2220.

Belavin et al. [3] these maps fall into homotopy classes corresponding to elements of the homotopy group, $\Pi_{d-1}(G)$. For most non-Abelian groups in four dimensional space time and U(1) in two dimensional spacetime, this homopoty group is Z. In these theories (the only ones we consider) the gauge fields integrated over in eq. (1) fall into discrete classes indexed by an integer ν running from $-\infty$ to $+\infty$ (we shall use the notation $[DA_\mu]_\nu$ to denote functional integration over the νth class). Thus there is actually a discrete infinity of funtional integrals and one must ask which, if any, is the "right" one.

One can clearly see what is going on by working in the gauge $A_0 = 0$ and requiring $F_{\mu\nu}$ to vanish outside a large, but finite, spacetime volume, V (this boundary condition is, of course, gauge invariant). The dynamical variables are now just the space components, A_i, for the vector potential and at large negative and positive times they must take on time independent vacuum values, $A_i(\boldsymbol{x}) = g^{-1}(\boldsymbol{x})\partial_i g(\boldsymbol{x})$. The topological quantum number, ν, associated with any particular Euclidean gauge field time history may be written as a gauge invariant volume integral

$$\begin{aligned}\nu &= \frac{1}{8\pi^2}\int \mathrm{d}^4x\ \mathrm{tr}\,(F_{\mu\nu}\tilde{F}^{\mu\nu}), \quad d=4\\ &= \frac{1}{4\pi}\int \mathrm{d}^2x\ \epsilon_{\mu\nu}F^{\mu\nu} \quad, \quad d=2.\end{aligned} \tag{2}$$

In both cases, the integrand is a total divergence and $A_0 = 0$ gauge ν may be rewritten as a surface integral, $\nu = n(t=+\infty) - n(t=-\infty)$, where

$$\begin{aligned}n &= \frac{1}{6\pi^2}\,\epsilon_{ijk}\int \mathrm{d}^3x\ \mathrm{tr}\,(A_i A_j A_k), d=4\\ &= \frac{1}{2\pi}\int \mathrm{d}x A_1 \ , \qquad d=2.\end{aligned} \tag{3}$$

With no loss of generality (we have the freedom of making time independent gauge transformations) we may choose $n(t=-\infty)$ to be an integer. Then since ν is integral the gauge vacuum configuration at $t=-\infty$ must also have integral winding number $n(+\infty) = n(-\infty) + \nu$.

Therefore we must admit the existence of a discrete infinity of vacuum states, $|n\rangle$, labelled by a winding number taking on integral values from $-\infty$ to $+\infty$. The interpretation of the multiplicity of Euclidean functional integrals corresponding to different ν-classes, is then straightforward:

$$\begin{aligned}\langle n|\exp(-Ht)|m\rangle \underset{t\to\infty}{\longrightarrow} &\int [DA_\mu \ldots]_{(n-m)}\\ &\times \exp\{-\int \mathrm{d}^dx[\mathcal{L}(A_\mu) + \mathcal{L}_{\mathrm{gf}}]\}\ .\end{aligned} \tag{4}$$

The functional integral over homotopy class ν describes a vacuum-to-vacuum transition in which the vacuum winding number changes by ν! Now the minimum action for $\nu = 0$ is zero (corresponding to $A_\mu \equiv 0$) so that in the WKB sense the $|n\rangle \to |n\rangle$ amplitude is O(1). In the $\nu \neq 0$ sectors the minimum action is in general non-zero – for $\nu = 1$ in four dimensions it corresponds to the Belavin et al. instanton [3], whose action is $8\pi^2/g^2$. Thus in the same WKB sense the $|n\rangle \to |n+1\rangle$ amplitude is $O(\exp(-8\pi^2/g^2)$. This is a typical "tunnelling" amplitude, vanishing exponentially for small coupling and unseen by standard perturbation theory. Indeed, perturbation treatments of gauge theories expand about $A_\mu = 0$ and pretend that the vacuum $|n=0\rangle$ is true vacuum. Because of vacuum tunnelling, this is completely wrong and causes perturbation theory to miss qualitatively significant effects.

What then is the true vacuum? A convenient way of constructing it is to consider the generators of time independent gauge transformations characterized by a gauge function $\lambda^a(\boldsymbol{x})$:

$$Q_\lambda = \int \mathrm{d}^{d-1}x[F^a_{0i}D_i\lambda^a + g\,J^a_0\lambda^a]$$

where D_i is the covariant derivative and J^a_0 is the gauge source of fields other than the gauge field itself. In order to satisfy Gauss' law, $D_iF^a_{0i} = gJ^a_0$, it is sufficient to restrict the state space by $Q_\lambda|\psi\rangle = 0$ for all gauge functions λ which vanish at infinity. In particular, all our vacuum states $|n\rangle$ are annihilated by such local gauge transformations. There also exist gauge functions which do not vanish at infinity and generate gauge transformations, T, which change the vacuum topology. One can easily construct a unitary T effecting such a non-local gauge transformation: $T = \exp(iG_\infty)$, with $G_\infty\ (2\pi/g)[E(\infty) + E(-\infty)]$ for the two-dimensional abelian theory or $G_\infty = (2\pi/g)\int \mathrm{d}^2S_iE^a_i\hat{x}^a$ for the four-dimensional non-Abelian theory, and T satisfies $T|n\rangle = |n+1\rangle$.

Since T is a gauge transformation, the hamiltonian commutes with it and energy eigenstates must be T eigenstates. Since T is unitary, its eigenvalues are $e^{i\theta}$, $0 \leqslant \theta \leqslant 2\pi$, and the eigenstates are $|\theta\rangle = \Sigma e^{in\theta}|n\rangle$. This diagonalization of H is obviously unaffected by including in $\mathcal{L}$ sources coupled to gauge invariant densities. Thus, each $|\theta\rangle$ vacuum is the ground state of an independent and in general physically inequivalent sector within which we may study the propagation of gauge invariant disturbances. Since the different θ-worlds do not communicate with each other, there is no a-priori way of deciding which world is the right one. It is gratifying that this multiplicity of worlds is known to exist in the Schwinger model, corresponding there to different values of background electric field [4].

Finally, we must express the functional integral, eq. (4), in θ basis:

$$\langle\theta'|\exp(-Ht)|\theta\rangle \xrightarrow[t\to\infty]{} \delta(\theta-\theta')I(\theta)$$

$$I(\theta) = \sum_\nu \exp(-i\nu\theta)$$

$$\times \int [DA_\mu \ldots]_\nu \exp\{-\int d^dx[\mathcal{L}(A_\mu \ldots) + \mathcal{L}_{gf}]\}$$

$$= \int [DA_\mu \ldots] \exp\{-\int d^dx[\mathcal{L}_{gf} + \mathcal{L}_\theta]\} \qquad (5)$$

where $\mathcal{L}_\theta = (i\theta/8\pi^2)\,\mathrm{tr}\,(F_{\mu\nu}\tilde{F}^{\mu\nu})$ for $d = 4$, $\mathcal{L}_\theta = (i\theta/4\pi)\,\epsilon_{\mu\nu}F_{\mu\nu}$ for $d = 2$ and in the second expression for $I(\theta)$ all gauge field topologies are summed over. $I(\theta)$ contains all possible information about physics in the θ-world and requires no further modification. The second form for $I(\theta)$ makes manifest one of the peculiar ways in which the θ-worlds differ from one another. In four dimensional pure Yang-Mills theory, re-expressed in Minkowski coordinates, the effective Lagrangian is $\mathrm{tr}\,[F_{\mu\nu}F^{\mu\nu} + (\theta/8\pi^2 F_{\mu\nu}\tilde{F}^{\mu\nu}]$. This clearly breaks P and T invariance (except for $\theta = 0$) and we must in general expect spontaneous breaking of space-time symmetries in all but a few special θ-worlds!

As a concrete illustration of these general remarks we should like to present an approximate evaluation of $I(\theta)$ in two-dimensional charged scalar electrodynamics. In the sector with $\nu = \pm 1$ the field configuration with minimum action is just the Nielson-Olesen vortex [5] in which there is a localized region of non-zero field of flux $\pm 2\pi/g$, radius μ^{-1} (μ is the heavy photon mass), arbitrary location and total action, S_0, proportional to μ^2/g^2. We shall construct the sectors with topological quantum number ν by superposing n_+ $\nu = +1$ vortices and n_- $\nu = -1$ vortices with $n_+ - n_- = \nu$, neglecting any interactions between vortices (since fields decrease exponentially this is not too bad for low vortex density, which turns out to mean small g). In this "dilute gas" approximation, the functional integral is

$$\langle\theta'|\exp(-Ht)|\theta\rangle \sim \delta(\theta'-\theta) \sum_{n_+,n_-=0}^{\infty} \exp\{-(n_+ + n_-)S_i\}$$

$$\times \frac{\exp\{i\theta(n_+ - n_-)\}}{n_+!\,n_-!}\left(\frac{V}{V_0}\right)^{n_+ + n_-} \qquad (6)$$

where the factors of V come from integrating over vortex locations and V_0 is a normalization factor which can be calculated from the quantum corrections to this basically semiclassical approximation. The sum is trivial and yields $\exp\,[2(V/V_0)e^{-S_0}\cos\theta)$. We have normalized the energy so that the naive perturbation theory vacuum energy is zero. By contrast, the θ vacua have an energy per unit volume equal to $-2V_0^{-1}\cos\theta\, e^{-S_0}$. Because $S_0 \propto 1/g^2$, this energy difference is a non-perturbative effect (a tunnelling effect) but potentially important nonetheless. Although the $\theta = 0$ vacuum has lowest energy (and no parity violation) we can't conclude that it is *the* vacuum since the other θ-vacua, though higher in energy are stable to gauge invariant perturbations. Having constructed a vacuum one can then calculate Green' functions of gauge invariant operators perturbatively. In the path integral this corresponds to performing ordinary perturbation theory about the appropriate classical solution for each topologically distinct sector and summing.

If we try the above sort of approximation on the non-abelian theory in four dimensions, there is a problem. The classical theory is scale invariant and the basic $\nu = 1$ solution (instanton) has an arbitrary scale parameter, λ, as well as an arbitrary position. The integration over λ need not diverge since scale invariance is broken by quantum corrections. Indeed, the renormalization group should tell us whether the integral converges at the short distance end. In the dilute gas approximation one finds for the vacuum energy density

$$\mathcal{E}_\theta = -2\cos\theta \int_0^\infty \lambda^3 \,\mathrm{d}\lambda \exp\left[-\frac{8\pi^2}{\bar{g}(\lambda/\mu)^2} + \mathcal{F}(\bar{g}(\lambda/\mu))\right] \tag{7}$$

where $\bar{g}(\lambda/\mu)$ is the usual effective coupling, normalized so that $\bar{g}(\lambda = \mu) = g$, μ is arbitrary, and $\mathcal{F}$ summarizing the effect of loop corrections, can be computed perturbatively. If the theory is asymptotically free and there are not too many quark multiplets, $\bar{g}$ can vanish rapidly enough for the integral to converge in the limit of large λ (small instanton size). This condition is met for any pure SU(N) gauge theory and for SU(3) with no more than *ten* flavors of quark.

On the other hand, in the limit of large instanton size, one is driven to large coupling (unless β has a small infrared fixed point) and the dilute gas approximation breaks down (instanton overlap and have long range interactions). Thus the attempt to construct the vacuum may run into an essential strong coupling problem because the quantum corrections to vacuum tunnelling will be large for large instanton size. In fact, there may not be a sensible way of perturbatively calculating even Green's functions of gauge invariant operators, no matter how small one makes g. This phenomenon is typical of a theory with no inherent mass scale which produces masses dynamically. If one sets the renormalization mass scale, μ, equal to some physical mass (e.g. $4\sqrt{\mathcal{E}_\theta}$), then g is determined (dimensional transmutation) and typically of order 1.

These problems should not, however, affect the standard applications of asymptotic freedom which rely on one's ability to compute operator product expansion coefficients at short distances. Precisely because of asymptotic freedom, vacuum tunnelling is suppressed at arbitrarily small scales and leading short distance behavior will agree with conventional calculations. There will, however, be calculated non-leading terms suppressed by powers of momentum, which reflect the mass scale set non-perturbatively by the tunnelling phenomenon.

The arguments presented above require some modification when massless fermions are present. We again confine non-zero $F_{\mu\nu}$ to a large but finite space-time volume, V, and again encounter a discrete infinity, $\{|n\rangle\}$, of vacuum states characterized by a vacuum gauge field with winding number n and a standard fermi vacuum with all negative energy states filled. In principle we must allow for transitions between vacuum, and evaluate $\langle n|\mathrm{e}^{-Ht}|m\rangle$ for general n and m. In fact, for massless quarks, $\langle n|\mathrm{e}^{-Ht}|m\rangle \propto \delta_{nm}$!

The reason for this is that, because of the anomaly, the *conserved* axial charge

$$Q_5 = \int \mathrm{d}\bar{x} J_c^5,$$

$$J_\mu^5 = \sum_{\text{flavor,color}} \bar{\psi}\gamma_\mu\gamma_5\psi - \mathrm{tr}\left\{\frac{g^2N}{32\pi^2}\epsilon_{\mu\nu\lambda\sigma}A_\nu(\partial_\lambda A_\sigma + \tfrac{2}{3}A_\lambda A_\sigma)\right\} \tag{8}$$

while invariant under local gauge transformations, is not invariant under global gauge transformations. In particular, one has $TQ_5T^{-1} = Q_5 - 2N$ where T is the global gauge transformation, introduced earlier, which changes gauge field winding number by one unit and N is the number of flavors. If the vacuum states of different topology are defined by $|n\rangle = T^n|0\rangle$, with $Q_5|0\rangle = 0$ one finds that $Q_5|n\rangle = 2N\cdot n|n\rangle$. However, Q_5 is *conserved*, so that it must be true that $\langle n|\mathrm{e}^{-Ht}|m\rangle \propto \delta_{n,m}$. In general, we must find $\langle n|\mathrm{e}^{-Ht}D_\nu|m\rangle \propto \delta_{n-m,\nu}$ where D_ν is any operator of chirality $2N\nu$ (D_ν may stand for multiple insertions of local operators at different times – all that matters is net chirality). Therefore, we may replace eq. (4) by

$$\langle n|\mathrm{e}^{-Ht}|m\rangle \to \delta_{nm}\int[DA_\mu\ldots]_0 \exp\{-\int \mathrm{d}^dx[\mathcal{L}(A_\mu\ldots) + \mathcal{L}_{\mathrm{gf}}]\}$$

and

$$\langle n|\mathrm{e}^{-Ht}D_\nu|m\rangle \to \delta_{n+\nu,m}\int[DA_\mu\ldots]_\nu D_\nu \times \exp\{-\int \mathrm{d}^dx[\mathcal{L}(A_\mu\ldots) + \mathcal{L}_{\mathrm{gf}}]\} \tag{10}$$

with the same meaning still attached to D_ν. The restriction on the topology of the gauge field histories would actually have emerged directly from a mindless application of eq. (4): Doing the fermion integrations for fixed A_μ yields $[\det(\not\partial - \not A)]^{+1}$. This determinant vanishes whenever $(\not\partial - \not A)$ has a zero eigenvalue. 't Hooft [6] has noted that if A_μ is taken equal to the $\nu = +1$ or $\nu = -1$ instanton there is a zero eigenvalue, and our argument is just telling us that whenever A_μ belongs to a $\nu \neq 0$ class, $(\not\partial - \not A)$ has a zero eigenvalue, eliminating the $\nu \neq 0$ sectors from the integration.

Though vacuum tunnelling is now suppressed, the

$|n\rangle$ vacua are not acceptable because they violate cluster decomposition for operators of non-zero chirality . Consider an operator D of chirality $2N$. The arguments of the preceeding paragraph show that $\langle n|D|n\rangle = 0$, $\langle n+1|D|n\rangle \neq 0$. Then $\langle n|D^+(x)D(y)|n\rangle$ will not vanish for large $|x-y|$ as required by cluster decomposition and the vanishing of the "vacuum" expectation $\langle n|D|n\rangle$: it obviously approached $\langle n|D^+|n+1\rangle\langle n+1|D|n\rangle$. The solution to this problem is obvious (it was solved in the Schwinger model years ago!): The proper vacuum states are the $|\theta\rangle$ vacua, in which basis the functional integrals have the form

$$\langle\theta'|e^{-Ht}|\theta\rangle \rightarrow \delta(\theta'-\theta)\int[DA_\mu\ldots]_{(0)}$$
$$\times \exp\{-\int d^dx[\mathcal{L}(A_\mu\ldots)+\mathcal{L}_{gf}]\}$$
$$\langle\theta'|e^{-Ht}D_\nu|\theta\rangle \rightarrow \delta(\theta'-\theta)\int[DA_\mu\ldots]_{(\nu)}$$
$$\times \exp\{-\int d^dx[\mathcal{L}(A_\mu\ldots)+\mathcal{L}_{gf}]\}D_\nu\,. \qquad (11)$$

The cluster problem is resolved by the non-vanishing vacuum expectation value of D in the true vacuum state. The fact that only one topological class of gauge field history contributes to each functional integral makes physical quantities have a trivial dependence on θ: The vacuum energy density, while non-zero, is independent of θ. The variation of the vacuum energy with respect to θ is just $\langle \mathrm{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu}\rangle$, the quantity whose non-zero value is the signal for P and T violation. In the massless fermion case, P and T appear not to be spontaneously violated and, indeed, all the $|\theta\rangle$ vacua are physically equivalent. Finally, the axial baryon number invariance of the original Lagrangian is violated by a vacuum expectation value of operators with non-zero chirality. It should be said that at this stage we have in the N-flavor case, only broken axial U(1) and *not* axial SU(N). Actually, since the axial charge rotates θ, a discrete subgroup of order $2N$ of U(1) is left unbroken, consisting of those elements which rotate θ by a multiple of 2π. There is no associated Goldstone boson because the conserved, but gauge-variant, U(1) charge takes one out of a given $|\theta\rangle$ sector ($e^{i\alpha Q_5}|\theta\rangle = |\theta+2N\alpha\rangle$) while $\mathrm{tr}(F\tilde{F})$, the divergence of the gauge invariant axial current, has non-vanishing matrix elements. That Q_5 causes transitions between different vacua is characteristic of the "vacuum seizing" mechanism postulated by Kogut and Susskind [2] while the non-vanishing of $F\tilde{F}$ in instanton solutions as a possible escape from the U(1) problem was noted by G. 't Hooft [7].

Although the presence of zero mass fermions suppresses vacuum tunnelling in the strict asymptotic sense, tunnelling does have a profound effect on the vacuum energy and other physically relevant quantities. When the vacuum tunnels, fermion pairs are produced. Although the pair must ultimately be absorbed by an anti-tunnelling, since the fermions are massless the pair may live for a long time and tunnelling occurs freely in intermediate states. To get some notion of what goes on it is instructive to attempt a crude calculation of the basic functional integrals of eq. (11) in the case of a single flavor.

We shall assume that the integral over A_μ is dominated by configurations of widely separated instantons (n_+ in number) and anti-instantons (n_- in number). To compute the vacuum energy we must set $n_+ = n_-$ (configurations with $\nu = 0$), sum over n_+ and integrate over instanton locations. We will ignore the integration over instanton sizes. For a given gauge field configuration, the integration over fermi fields yields $\det(\not{\partial} - \not{A})$. This determinant must also be approximated.

Now, as 't Hooft has pointed out [7], individual instantons have a zero-energy eigenfunction $\psi_0{}^{\pm}(x, x_\pm)$ ($x_\pm$ is the instanton location and the ± label distinguishes instanton from anti-instanton). Since the interesting physical effects arise precisely from these zero energy solutions, we shall compute the determinant of $(\not{\partial} - \not{A})$ in the subspace spanned by the $2n_+$ functions $\psi_0{}^+(x, x_i^+)$, $\psi_0^-(x, x_i^-)$. In the widely separated instanton approximation, these functions are orthonormal and one has to compute the determinant of the matrix

$$M_{ij} = (\psi^*_{0,i}|(\not{\partial} - \not{A})|\psi^-_{0,j}).$$

In this approximation A_μ differs from a gauge transformation only in the neighborhood of each instanton and M_{ij} can be approximated by

$$M_{ij} \approx (\phi^+_{0,i}|(-\not{\partial})^{-1}|\phi^-_{0,j})$$

where $\phi_0 = \not{\partial}\psi_0$. The γ_5 structure of the ψ_0 forbids i and j to be both instantons or both anti-instantons. The ψ_0 fall off at large distances exactly like the free fermi propagator, which is why we introduce the ϕ_0 's which are localized as well as the instanton itself. The cycle expansion of the determinant then gives a sum of terms with a graphical interpretation in terms of closed fermion loops. For instance, if $n_+ = n_- = 1$ we get

$$\mathrm{Det} \approx (\phi_0^+ |(\not{\partial})^{-1}| \phi_0^-)(\phi_0^- |(\not{\partial})^{-1}| \phi_0^+) ,$$

which has the obvious interpretation of massless quark propagators connecting non-local vertices, $V^\pm = \phi_{0\pm}^*(x)\phi_{0\pm}(x')$, associated with the instantions. The functional integral weights each vertex with a factor proportional to $\mathrm{e}^{-S_{\mathrm{cl}}}$, where S_{cl} is the instanton classical action. The γ_5 structure of $\phi_0^\pm$ is such that $V^\pm$ is like $\bar{\psi}(1 \pm \gamma_5)\psi$ in its Dirac matrix structure, which is to say that $V^\pm$ looks like a non-local, or momentum-dependent mass term. Summing over numbers and locations of instantons simply completes the vacuum fermion loop analogy by providing all possible insertions of the pseudo mass terms, $V^\pm$, on massless quark loops. Then calculations of physical quantities proceed in a perfectly conventional way so long as we remember to add the mass term $\sim \mathrm{e}^{-S_{\mathrm{cl}}}(V_+ + V_-)$ to the massless quark propagator. Anything which directly depends on the mass term, such as the vacuum expectation of $\bar{\psi}\psi$, will be proportional to $\mathrm{e}^{-S_{\mathrm{cl}}}$ and will have only the by now familiar dependence on g characteristic of a tunnelling process. In the "scale invariant" four dimensional theory $\langle\bar{\psi}\psi\rangle \neq 0$ implies spontaneous generation of mass and dimensional transmutation as before.

If the fermion is given a bare mass tunnelling is allowed and one is driven directly to a θ vacuum (whose energy depends on θ). The limit of zero mass is smooth. If the bare mass is small compared to the spontaneously generated mass, it acts as a small perturbation on the $m_0 = 0$ theory.

If there are N flavors, the above discussion is modified in an important way: the effective instanton-quark interaction is no longer billinear, but $2N$-linear. Indeed, the instanton (anti-instanton) vertex has the structure $V_+ = \Pi_{i=1}^N \bar{\Psi}_i(1+\gamma_5)\psi_i$ ($V_- = \Pi_{i=1}^N \bar{\Psi}_i(1-\gamma_5)\psi_i$). As a result, summing over instantons in the dilute gas approximation will not just produce a mass term in the quark propagator, but does something more complicated. To produce a quark mass term, one must break the global chiral SU(N), while we have argued that the vacuum tunnelling phenomenon is only guaranteed to violate the chiral U(1) symmetry. On the other hand, the effective interactions between quarks of different helicity generated by the instanton provide new ways of identifying sums of graphs which can lead to the desired symmetry breakdown. We have constructed a simple Hartree–Fock type argument for $N = 2$ which has a chance of being correct in a weak coupling theory and which seems, on superificial examination, to generate quark masses. The inevitable Goldstone bosons arise in this case from iterated bubble graphs generated by the four-fermion interactions, $V_\pm$. We do not wish to make too much of these crude arguments other than to suggest that the new interactions generated by vacuum tunnelling are likely to play a key role in generation of quark masses and Goldstone bosons.

In terms of the picture presented here Polyakov's ideas about confinement appear as follows. For an isolated quark located at $\boldsymbol{x}$ the tunnelling amplitude $|n, \boldsymbol{x}\rangle \to |n+1, \boldsymbol{x}\rangle$ will be reduced relative to the vacuum to vacuum amplitude. A quark state will then have more energy than the vacuum, as it should. In the dilute gas approximation the energy difference is proportional to the integral over all instantons which overlap the quark. Integrating over instanton locations and then over the scale size λ^{-1} leads to an integral which tends to diverge at small λ. For the large instantons (small λ), however, the dilute gas approximation is not valid, and one is again confronted with a strong coupling problem.

Obviously, much remains to be done to fully exploit the phenomena we have found. The major difficulty, of course, is that the theory we are really interested in, quantum chromodynamics, is basically a strong coupling theory and reliable calculations are difficult, if not impossible. However, one may hope that a new understanding of the qualitative physics will suggest new methods of calculation. We are especially encouraged by the appearance, already in semiclassical approximations, of a vacuum that breaks chiral symmetry and sets a dynamical mass scale. We are also intrigued by the natural appearance of spontaneous violation of P and T invariance but have so far not seen how to understand why these effects are small in the real world or how to exploit them to explain observed violations of these symmetries. Perhaps super-unified theories will shed some light on these questions.

One of us (D.J. Gross) would like to acknowledge V.N. Gribov for stimulating conversations and in par-

ticular for the suggestion that the Euclidean solitons might be relevant to the structure of the vacuum.

References

[1] A.M. Polyakov, Phys. Lett. 59B (1975) 82.
[2] J. Kogut and L. Susskind, Phys. Rev. D11 (1975) 3594.
[3] A.A. Belavin, A.M. Polyakov, A.S. Schwartz and Yu. S. Tyupkin, Phys. Lett. 59B (1975) 85.
[4] J. Lowenstein and A. Swieca, Ann. Phys. 68 (1971) 172. S. Coleman, Harvard preprint (1975).
[5] H. Nielsen and P. Olesen, Nucl. Phys. B61 (1973) 45.
[6] T.S. Bell and R. Jackiw, Nuovo Cimento 60 (1969) 47; S. Adler, Phys. Rev. 177 (1969) 2426.
[7] G. 't Hooft, Harvard preprint (1976).