

COSMOLOGY AND BROKEN DISCRETE SYMMETRY

John PRESKILL*

California Institute of Technology, Pasadena, CA 91125, USA

Sandip P. TRIVEDI**

*School of Natural Sciences, Institute for Advanced Study, Olden Lane,
Princeton, NJ 08540, USA*

Frank WILCZEK***

*School of Natural Sciences, Institute for Advanced Study, Olden Lane,
Princeton, NJ 08540, USA*

Mark B. WISE*

California Institute of Technology, Pasadena, CA 91125, USA

Received 2 April 1991

It is argued that under various circumstances spontaneously broken discrete symmetry, which is commonly thought to lead to grave cosmological problems, need not do so. A particularly interesting example is the two-doublet version of the standard model. This extended model avoids unacceptable flavor-changing neutral currents in a natural way only if an appropriate discrete symmetry is postulated. Here it is shown that this discrete symmetry is anomalous. As a result, when it is spontaneously broken it does *not* produce stable domain walls (which would be problematic for cosmology). Instead an interesting cosmological scenario results, which incorporates substantial deviations from thermal equilibrium at the QCD scale. If axions are used to address the strong *CP* problem, then the analysis is altered. Depending on the Peccei–Quinn charge assignments, domain walls may be either stable or nonexistent.

1. Problems of discrete symmetry

One of the earliest applications of cosmological arguments to fundamental particle physics – and still among the most significant – is the observation by

* Research supported in part by DOE grant DEAC-03-81-ER40050

** Research supported in part by NSF grant NSF-PHY-8620266

*** Research supported in part by DOE grant DE-FG02-90ER40542

Zel'dovich et al. [1] that spontaneously broken discrete symmetry, which may be an attractive possibility in building models of particle physics, can lead to grave difficulties in the context of cosmology. The reason for the difficulties is very simple. Spontaneously broken discrete symmetry implies the existence of several degenerate ground states. In the big bang, there is a natural limitation to the homogeneity of the phase transition at which the spontaneous symmetry breaking first occurs. Causally disconnected spatial regions will have no special tendency to make the same choice of degenerate ground state. And so at a later stage in the evolution, when the homogeneous ground state is approached everywhere locally, the different ground-state regions will be separated by stable domain walls. Insofar as previously causally disconnected regions are always newly entering the horizon, we must expect that at least one domain wall of roughly the dimension of the horizon will exist. The energy per unit area of a domain wall is set by a microphysical parameter; call it ε . Then the energy in the domain wall will be at least of order $R^2\varepsilon$, where R is the size of the horizon (radius of the universe), corresponding to an energy density $\rho_{\text{dw}} \sim \varepsilon/R$. This decrease in energy density is much slower than one finds for ordinary radiation or nonrelativistic matter. Thus stable domain walls will quickly come to dominate the mass of the universe, and continue to do so in gross disagreement with observation.

This conclusion is rather disappointing, since spontaneously broken discrete symmetries are an important ingredient in many interesting models of particle physics. Here we shall argue that there are quite natural and appealing ways out. One possibility is that the discrete symmetry is *anomalous*. In this case the "ground states" which appear at the classical level and to all orders of perturbation theory to be degenerate are found instead to be separated by a finite energy difference when they are compared nonperturbatively in the quantized theory. We shall see that this possibility occurs in perhaps the simplest interesting extension of the minimal standard model, namely its extension to include two Higgs doublets. In this context, a discrete symmetry is required in order to banish potential unacceptable flavor-changing neutral currents in a natural way [2]. This symmetry turns out to be anomalous. As a result the specter of a universe dominated by stable domain walls is exorcised, replaced with an apparently benign and potentially quite interesting cosmological scenario.

The anomalous symmetry of the two-doublet model can be considered as relating $\theta = 0$ and $\theta = \pi$ phases of QCD. This is one indication (among many) that this symmetry is closely tied up with the strong *CP* problem. It therefore behooves us to inquire how our considerations are affected by the most likely solution of this problem, involving Peccei–Quinn (PQ) symmetry [3] and axions [4]. As might be expected (since the main point of PQ symmetry is to make the θ angle of QCD a dynamical variable), the effects are profound. We find that the domain walls may again become stable, or may cease to exist altogether, depending on how the PQ symmetry is implemented.

2. Discrete symmetry of the two-doublet standard model

Although some of our considerations may be of more general interest, we shall in this paper focus on the two-doublet extension of the minimal standard model. (More specifically, we consider the version of this model in which both doublets play a role in giving mass to quarks and leptons, and tree-level flavor-changing neutral currents are avoided in a natural way.) This model is interesting for several reasons. It is the simplest model in which the mechanism of mass generation is not universal. It also occurs as a “sub-model” in low-energy supersymmetry and in attempts to introduce significant symmetry structure in the flavor sector.

The mass terms for the quarks are taken to be of the form

$$L_m = \bar{Q}_\alpha^i U^j g_{ij} \phi_U^\alpha + \bar{Q}_\alpha^i D^j h_{ij} \phi_{D\beta}^* \epsilon^{\alpha\beta} + \text{h.c.} \quad (2.1)$$

Here the Greek indices represent weak SU(2) quantum numbers, the Latin indices represent flavor quantum numbers, and g and h are matrices of (generally, complex) coupling constants. Color SU(3) indices are suppressed. Q is a left-handed weak SU(2) doublet of fermions and U, D are right-handed singlets. It is very important that each type of quark receive mass from just *one* Higgs field. Otherwise, generically one would encounter flavor-changing neutral-current processes at unacceptable rates. In order to insure this feature of (2.1) in a natural way, one postulates the existence of the discrete symmetries

$$\phi_U \rightarrow -\phi_U, \quad U \rightarrow -U, \quad (2.2)$$

or

$$\phi_D \rightarrow -\phi_D, \quad D \rightarrow -D. \quad (2.3)$$

(Only one of these need be postulated, since both together generate a transformation which is a product of $(-1)^F$ and a hypercharge rotation*.) These symmetries do enforce the structure (2.1) and forbid the dangerous neutral current processes in a natural way.

However, although these discrete symmetries are very desirable from the point of view of particle physics, they pose dangers for cosmology. The symmetries (2.2) and (2.3) are spontaneously violated when the fields ϕ_U, ϕ_D acquire vacuum expectation values and the quarks acquire mass. Thus the theory would appear to contain the possibility of stable domain walls separating the degenerate vacua. In the evolution of the Universe according to big bang cosmology such domain walls would be formed at the phase transition where the ϕ 's acquire vacuum expectation

* This is true even when leptons are included. Each charged lepton is Yukawa coupled to either ϕ_U or ϕ_D . The combination of (2.2) and (2.3) is a symmetry of the lepton sector if, under the combined transformation, lepton doublets are invariant and lepton singlets change sign.

values. They would produce marked anisotropies in the Universe and would even eventually come to dominate the mass density of the Universe, all contrary to observation.

3. Forms of discrete symmetry

Can this negative conclusion be avoided? One should consider the question carefully, since the two-doublet model in question is otherwise among the most plausible foils to the minimal model; furthermore similar questions regarding the status of spontaneously broken discrete symmetries will arise in many other models.

Recently the possibility of gauged discrete symmetries has been widely discussed [5]. If a discrete symmetry is gauged, it would appear that the associated domain wall has lost its *raison d'être*. For gauge symmetry indicates a redundancy in the description of the states, and in this context would imply that the two states on the opposite sides of the wall were really the same state, whose two descriptions on the opposite sides are connected by a gauge transformation.

This "solution" sounds suspiciously metaphysical, and begs the question of a concrete physical implementation. Despite our decree that the discrete symmetry is gauged, the same field configurations that previously described the stable domain wall stubbornly remain solutions to the classical field equations.

The non-trivial physics associated with gauge discrete symmetry become manifest only when one considers string configurations which carry non-trivial flux dual to the discrete charge. For these strings to arise we must consider a situation in which the discrete symmetry is embedded in a continuous gauge symmetry. The evolution of the universe would then involve breaking the continuous gauge symmetry to its discrete subgroup at a scale V followed by the breaking of the discrete gauge symmetry at the scale v . The first transition would give rise to a string network and the second to domain walls. But a careful analysis shows that the domain walls formed would be bounded by strings. The behavior of these domain walls bounded by strings is quite different from a network of domain walls alone [6–8]. Tension in the domain walls tends to pull the strings together as they vibrate. Unless the intercommutation rate is very small the walls then rapidly cut themselves into ribbons which then decay by emitting gravitational radiation. Thus the cosmological disaster mentioned earlier is averted.

Of course, if the universe goes through an inflationary phase sometime between the two phase transitions, there would effectively be no strings left when the discrete symmetry got broken. The true dynamical significance of the gauge symmetry in this context is that it allows a process whereby the wall is punctured and destroyed by virtual string loops.

We can estimate the rate of this nucleation process using a simple argument based on the uncertainty principle [6, 8]. Let the energy of the wall per unit area be

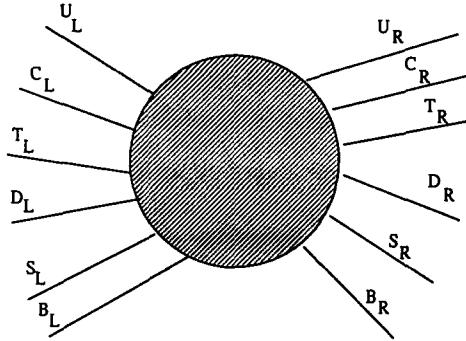


Fig. 1. The effective instanton vertex for QCD has one left-handed quark of each flavor coming in, and one right-handed quark of each flavor coming out.

denoted v^3 . (Thus crudely speaking v is the scale at which the discrete symmetry is broken.) Let the tension of the relevant string be V^2 . (Thus crudely speaking V is the scale at which the continuous gauge symmetry from which the discrete one derives is broken.) In order for the puncture to expand classically it must achieve a certain minimum size $L_c \geq V^2/v^3$, so that the energy gain from destroying portions of the domain wall can exceed the energy associated with the increased length of the string. To produce strings of this size by tunneling, one must consider nucleation of a tiny string and its classically forbidden evolution to critical size. The action for this classically forbidden process is generically of order $\int_0^{L_c} V^2 dL \sim (V/v)^6$, and the rate is therefore of order $\exp[-(V/v)^6]$. This is a disappointing rate in the context of our present problem, where a reasonable expectation might be $v \sim 10^2$ GeV and $V \sim 10^{15}$ GeV. (These processes can occur in the case of walls bounded by strings too but this estimate shows they would have a negligible effect.) Thus in case the strings get inflated away we are left with practically stable domain walls and their unacceptable consequences.

Actually, even if we wanted to embed the discrete symmetry (2.2) in a gauge symmetry we would not be able to do so, because this symmetry is *anomalous*. The notion of anomalous discrete symmetry has not been discussed explicitly before* (as far as we know), but it is an important one. We may visualize the anomaly quite concretely by considering the effective instanton vertex for QCD (fig. 1). Three right-handed charge $2/3$ quarks emerge from this vertex. Since three is an odd number, the vertex is not invariant but rather changes sign under the transformation (2.2). Because the putative discrete symmetry is not one in fact – it is a symmetry of the classical theory, but inevitably destroyed upon quantization – there is actually a difference in the physics generally, and in the energy density in

* A very recent exception has been brought to our attention [20].

particular, between the two putative degenerate ground states. These differences have some interesting interpretations and ramifications, to which we now turn.

4. Particle physics of anomalous discrete symmetry

If we consider temperatures or energies so low that excitation of the massive quanta associated with ϕ_U and ϕ_D can be neglected and the only relevant degrees of freedom are those of QCD, a very simple interpretation of the two quasi-vacua related by (2.2) may be given. They differ precisely in the sign of the amplitude of the instanton process – they simply represent the ground states, whose QCD θ angle differs by π ! Since P and T are observed to be conserved in the strong interactions, we must demand that the ground states of choice are $\theta = 0$ and $\theta = \pi$ very nearly.

Because the difference between the two states on opposite sides of the domain wall reduces to this well studied distinction, we can draw on previous work to determine some of the key parameters. The distinction between $\theta = 0$ and $\theta = \pi$ can be cast as a difference in the sign of the u -quark mass. The effect of light quark masses on the pseudoscalar meson masses allows one to determine their ratios; it is found that all the ratios are positive, corresponding to $\theta = 0$. One may also calculate, in chiral perturbation theory, the dependence of the vacuum energy on light quark masses, and thus extract the energy difference between $\theta = 0$ and $\theta = \pi$. At zero temperature the difference in energy density between the two states is related to quark masses and chiral symmetry parameters as

$$\epsilon = f_\pi^2 m_\pi^2 m_u / (m_u + m_d). \quad (4.1)$$

Also at temperatures well above 1 GeV (but still well below the mass of the ϕ -quanta), where QCD instantons may be treated as a dilute gas, it is possible to estimate the energy difference analytically. The result is [9]

$$\epsilon = 8 \times 10^{-4} \Lambda_{\text{QCD}} m_u m_d m_s \left(\frac{\Lambda_{\text{QCD}}}{\pi T} \right)^8 \left[9 \ln \left(\frac{\pi T}{\Lambda_{\text{QCD}}} \right) \right]^6. \quad (4.2)$$

The most notable feature of (4.2) is that the energy difference falls very rapidly at high temperature. This reflects the weak dependence of the free energy of high-temperature QCD on the θ -parameter, due to the suppression of large instantons by screening (and of small ones by asymptotic freedom).

5. Cosmology of anomalous discrete symmetry

The distinctive cosmological consequences of our broken anomalous discrete symmetry follow directly from its peculiar particle physics as outlined above. We shall first sketch its broad outline, then attempt to fill in some details.

The main feature is that this symmetry is *spontaneously* broken by a large amount and *intrinsically* broken by a small amount. (One would normally say it is strongly spontaneously broken and weakly intrinsically broken – however that terminology would be terribly confusing here, because the strong spontaneous breaking is associated with the weak interactions while the weak intrinsic breaking is due to the strong interactions!) The symmetry breaking sets in at the weak scale, and the energy barriers between the nearly degenerate vacua are generically also of this magnitude. The energy difference between the vacua, on the contrary, is set by the strong scale – and actually, as we have seen, further suppressed by light quark masses. Furthermore, this energy difference depends sharply on the temperature, and vanishes rapidly as this temperature exceeds the strong scale. Thus we expect that at the temperatures where the symmetry spontaneously breaks the intrinsic energy difference between the two nearly degenerate vacua has negligible dynamical import. At and below these temperatures, until the strong interactions kick in, the cosmology will develop as if it were heading towards a domain-wall dominated universe.

However, at the strong scale the energy difference will cease to be negligible. It creates a pressure difference driving the domain walls into the false vacuum regions, and ultimately squeezing them out. One must remember, though, that the energy difference is unusually small. Since the pressure it exerts must compete with the surface tension and viscosity of the walls, and must contend with their large inertia, the process of establishing true vacuum everywhere can be drawn out over time and can lead to large-scale inhomogeneities at intermediate stages.

Now let us try to analyze these events in some more detail. What does the universe look like immediately after the phase transition? This question can be answered using the techniques of percolation theory. The two Higgs fields acquire vacuum expectation values which are characterized by a correlation length ξ . We can then crudely picture the universe as a cubic lattice with links of length ξ and sites which have an equal probability of being in either of the two phases [10, 11]. In three dimensions the percolation threshold p_c is less than 0.5 [12]. This means that infinite-sized connected clusters of both phases will appear in the universe. In fact, most of the volume of either phase appears in one infinite cluster. Since the domain walls mark the two phases, the system of domain walls will be dominated by one infinitely long domain wall of very complicated geometry. In addition there will be some closed domain walls of finite size most of which have dimensions of order ξ .

How will such a system evolve? As eq. (4.2) indicates, initially the energy difference between the two phases is negligible. Consequently these walls will move under the influence of the wall tension σ and the viscosity of the surrounding medium [10]. One source of viscosity is that some quarks and also possibly leptons have opposite masses on the two sides of the wall [13]. This gives them a non-zero probability to reflect off the wall, and results in a viscous force which

opposes the motion of the wall. A simple calculation shows that for an infinitely thin plane wall the probability of reflection is $(m_q/E)^2$ at normal incidence, where m_q is the mass of the quark and E its energy. To get the total viscosity we would have to sum over the contributions of all the quarks and leptons which are relativistic and which change their mass when going across the domain wall. Besides this there is another potentially interesting contribution to the viscosity coming from charged fermion zero modes which make the wall superconducting. We shall not attempt here to treat this rich dynamics in detail, and for simplicity in the subsequent discussion will crudely model the pressure due to viscosity on a wall moving with velocity v as

$$P_v = T^4 v. \quad (5.1)$$

(This is a reasonable order of magnitude estimate if the wall is nearly impenetrable to at least one species of relativistic radiation. The actual viscosity could be much smaller if every species penetrates the wall easily.)

In the absence of viscosity we might have expected the walls to straighten out on the scale of the horizon. But with the viscosity playing an important role only irregularities on much smaller scales get straightened out. Curvature on the scale R leads to motion with terminal velocity v given by

$$P_v = \sigma/R. \quad (5.2)$$

Here σ is the wall tension, which is about the same order of magnitude as the mass per unit area of the wall. Putting in our crude estimate for the viscous force, we find

$$T^4 v = \sigma/R. \quad (5.3)$$

Thus the time for motion on the scale R which is given by

$$t_d = R/v \quad (5.4)$$

is

$$t_d = R^2 T^4 / \sigma. \quad (5.5)$$

It is reasonable to expect that if this time is less than the Hubble time irregularities on the scale R will be smoothed out. This sets a scale $R_s(T)$ given by

$$R_s(T)^2 T^4 / \sigma = M_{\text{Pl}} / T^2,$$

i.e.

$$R_s(T) = \sqrt{\sigma M_{\text{Pl}}} / T^3, \quad (5.6)$$

on which the walls are smooth.

We mentioned earlier that immediately after the electroweak phase transition the walls are flat on a scale characterized by a correlation length ξ of the vev's of the Higgs fields. The precise value of ξ depends on the detailed nature of the phase transition. Causality limits it to be smaller than the horizon size but in fact in case of a second order phase transition it can considerably smaller and of the order of T^{-1} . The evolution of the domain walls depends on ξ . If ξ is of the order of the horizon size at the weak scale it is much larger than $R_s(T)$ even when the temperature falls to the QCD scale. So the viscosity would prevent the walls from straightening out any further and we would expect the walls to conformally stretch with the expansion. Thus the scale $R_s(T)$ on which they are flat is actually given by

$$R_s(T) \approx M_{\text{pl}} / (T_w T) \tag{5.7}$$

and the energy density in the walls goes as

$$\rho_{\text{wall}} \approx \sigma / R_s(T). \tag{5.8}$$

On the other hand, if ξ is approximately T^{-1} and therefore much smaller than the horizon, then as we have seen above the walls will straighten out somewhat and $R_s(T)$ will be given by eq. (5.6), so that

$$\rho_{\text{wall}} \approx \sqrt{\sigma / M_{\text{pl}}} T^3. \tag{5.9}$$

In either case $\rho_w / \rho_{\text{rad}} \ll 1$ if the domain walls collapse around the QCD scale (see below).

It is difficult to determine exactly when the domain walls collapse. But once the energy difference between the two phases is large enough to result in relativistic terminal velocities we would expect the collapse to occur quickly. Equating the viscosity with the energy difference shows this happens when

$$\epsilon \approx T^4. \tag{5.10}$$

To estimate ϵ , we go back to eqs. (4.1) and (4.2). Strictly speaking it is of course not correct to use eq. (4.2) for temperatures close to Λ_{QCD} , but it can be used for our present purpose of rough estimation. Setting $T = \Lambda_{\text{QCD}}$, we find

$$\epsilon / (\Lambda_{\text{QCD}})^4 \approx 10^{-4}, \tag{5.11}$$

while eq. (4.1) shows that at zero temperature

$$\epsilon / (\Lambda_{\text{QCD}})^4 \approx 10^{-2}. \tag{5.12}$$

(In making these estimates we have taken $\Lambda_{\text{QCD}} = 200 \text{ MeV}$.) Thus the condition

in eq. (5.10) should be met soon after the chiral phase transition, and the domain walls never come close to dominating the energy density of the universe.

Can this process give an appreciable background of gravitational radiation? The energy density of the gravitational waves produced depends crucially on two parameters, the energy difference between the two phases when the domain walls collapse and the size of the bubbles of the higher energy phase. An order of magnitude estimate [14] is that a bubble of size R generates fluctuations in the metric with amplitude

$$h \approx G\epsilon R^2.$$

Since

$$\rho_{\text{gw}} \approx 1/Gh^2\omega^2, \quad (5.13)$$

taking

$$\omega \approx 2\pi/R$$

gives

$$\rho_{\text{gw}} \approx (2\pi)^2 G\epsilon^2 R^2. \quad (5.14)$$

Our discussion above indicates that these parameters depend on the details of the model (through the initial correlation ξ) and of the subsequent complicated evolution. This makes them difficult to estimate. We content ourselves with a rough estimate and in the discussion below take $\epsilon \sim \Lambda_{\text{QCD}}^4$ and R to be $R_s(\Lambda_{\text{QCD}})$. Then focusing on the more promising scenario, in which $R_s(T)$ is given by eq. (5.7), and denoting the horizon at the QCD scale by R_{H} we have

$$\rho_{\text{gw}} \approx 4\pi^2 G (\Lambda_{\text{QCD}}^4)^2 R_{\text{H}}^2 (R_s(\Lambda_{\text{QCD}})/R_{\text{H}})^2, \quad (5.15)$$

which implies that

$$\frac{\rho_{\text{gw}}}{\rho_{\text{rad}}} \approx \frac{10}{g_*} \left(\frac{R_s(\Lambda_{\text{QCD}})}{R_{\text{H}}} \right)^2. \quad (5.16)$$

Here g_* , the number of effectively massless degrees of freedom at the QCD scale, is about 65. Since

$$R_s(\Lambda_{\text{QCD}})/R_{\text{H}} \approx 2 \times 10^{-3}, \quad (5.17)$$

we find that the energy density in gravitational waves today corresponds to

$$\Omega_{\text{gw}} \approx 10^{-11}. \quad (5.18)$$

The present day frequency of these waves would be about 10^{-5} Hz, that is, about

the order of a (day)⁻¹. This crude estimate indicates that the gravitational radiation is much too small to be measured by pulsar timing. It also shows that the expected radiation depends sensitively on unknown details of the microscopic model and of the dynamical evolution. It is conceivable that a more careful analysis would show that for some values of the model parameters, the gravitational radiation is detectable in the proposed Lagos experiment [15].

This concludes our discussion of the cosmological implications of spontaneously broken anomalous discrete symmetry. Clearly a lot more work would be necessary, both in refining the theory and in determining the model parameters, to derive a convincing quantitative estimate of the gravitational radiation. However we believe that the main qualitative conclusion – that the anomalous discrete symmetry is not inevitably a cosmological catastrophe – is secure.

6. Incorporation of Peccei–Quinn symmetry

It is easy to anticipate that the incorporation of Peccei–Quinn symmetry will alter the analysis drastically. After all, the difference between the two quasi-vacua is precisely that they correspond to different values of the QCD θ -parameter, whereas the whole point of Peccei–Quinn symmetry is to make θ into a dynamical variable capable of – and desirous of – relaxing to zero from any initial value.

The precise realization of this anticipation is as follows. By definition, Peccei–Quinn symmetry allows one to compensate any given value of the θ -term by a transformation which otherwise leaves the lagrangian invariant. Thus the failure of (2.2) and (2.3) to be an exact symmetry, which arises solely from the fact that it changes θ by π , can be compensated by combining it with an appropriate Peccei–Quinn transformation. The combination will then be a true (non-anomalous) symmetry.

Depending on its nature, the combined non-anomalous symmetry may or may not be problematic. If it is discrete and spontaneously broken, we arrive back at the difficulties pointed out in ref. [1]. This is the case in one large class of axion models, the Kim [16] models*. In these models the Peccei–Quinn charge of the light quarks and Higgs field is zero; the Peccei–Quinn transformations act trivially on them. (The Kim mechanism is well illustrated by a model incorporating a complex $SU(3) \times SU(2) \times U(1)$ singlet scalar ϕ coupled to a heavy color triplet quark Q according to

$$\Delta L = g\bar{Q}_L Q_R \phi + \text{h.c.} \tag{6.1}$$

and fully invariant – at the classical level – under the PQ transformation $Q_R \rightarrow e^{i\alpha} Q_R$, $\phi \rightarrow e^{-i\alpha} \phi$. This transformation is anomalous, and plays its appropriate

* That is, domain walls at the weak scale are problematic in a Kim model *with* two light Higgs doublets. But the Kim mechanism does not *require* that there be two light doublets.

role in resolving the strong CP problem within the PQ scheme. If ϕ develops a large vacuum expectation value and Q a large mass, then one will have an axion corresponding to the phase of ϕ . This axion will couple to light matter through its anomalous couplings to gauge fields, but has no other more intrinsic couplings.) By combining (2.2) with the appropriate PQ transformation, one forms a nonanomalous discrete symmetry, which is spontaneously broken when ϕ_U and ϕ_D condense. This produces cosmological disaster. As in the two-doublet model we analyzed earlier, domain walls form at the weak transition. However, whereas in that case an energy difference developed between the quasi-vacua on opposite sides of the wall, in the present case the axion field will relax (due to Hubble damping) to produce $\theta = 0$ effectively on both sides, and the wall will feel no pressure. Each domain wall will be sheathed by an axion domain wall across which the phase of ϕ changes by π over roughly an axion Compton wavelength, but the bulk physics and the overall energy balance are little affected by this dressing.

The situation can be quite different for the Dine–Fischler–Srednicki–Zhitnitskii (DFSZ) models [17] in which ϕ_U, ϕ_D and the quarks carry non-zero PQ charge*. For in this case the combined symmetry can turn out to be essentially unbroken, or to be more precise harmlessly related by hypercharge to an unbroken symmetry. An explicit example should make this clearer. Suppose again that we have in addition to ϕ_U and ϕ_D the complex singlet scalar ϕ , but now including the cross-coupling

$$\Delta L_{\text{DFSZ}} = \phi \phi_U^\dagger \phi_D + \text{h.c.} \quad (6.2)$$

The term in eq. (6.2), as well as the couplings (2.1), will be invariant under the PQ transformation

$$\begin{aligned} \phi_U &\rightarrow e^{i\alpha} \phi_U, & \phi_D &\rightarrow e^{-i\alpha} \phi_D, & \phi &\rightarrow e^{2i\alpha} \phi, \\ U &\rightarrow e^{-i\alpha} U, & D &\rightarrow e^{-i\alpha} D. \end{aligned} \quad (6.3)$$

Now it is easy to see that under (6.3) the instanton vertex gets multiplied by $e^{-6i\alpha}$. Thus (6.3) with $\alpha = \pi/2$ combined with (2.2) gives a non-anomalous discrete symmetry. (The field ϕ must also change sign under (2.2) so that (2.2) is a symmetry of ΔL_{DFSZ} .) Under this combined symmetry both ϕ_U and ϕ_D get multiplied by the same phase $-i$. Thus this phase can be absorbed into a hypercharge transformation, leaving a residual transformation which reduces to multiplication by a common phase on all the quark fields. Finally this last symmetry is unbroken; it is a baryon number phase rotation. (When leptons are included, a piece proportional to the lepton number phase also appears.)

The cosmology of the scheme is quite simple too. Indeed once PQ symmetry is broken one finds that (6.2) induces a term which violates our original discrete

* We consider here DFSZ models with two light Higgs doublets, although the DFSZ mechanism does not require both doublets to be light.

symmetry (2.2). Thus at the weak scale the discrete symmetry (2.2) appears to be violated explicitly, and there is no question of domain walls. (Note that once ϕ acquires a vacuum expectation value ϕ_U and ϕ_D have the same quantum numbers, and it may not be apparent why the dangerous mass terms coupling ϕ_U to D and ϕ_D to U cannot appear. In fact they can, but only as finite – and generically small – radiative corrections. The original symmetry, though broken, plays its essential role in forbidding the strictly renormalizable mass-generating interactions with dimensionless coupling constants, which in the absence of symmetry would be incalculable subtraction terms.)

Actually this model is not quite adequate as it stands, because of a problem somewhat peripheral to our interests here. This is the problem that the PQ symmetry itself is non-anomalous not only at $\alpha = 0$ but also at $\alpha = k \cdot 2\pi/6$ for any integer k . The extra non-anomalous discrete symmetries are of course broken when ϕ_U and ϕ_D acquire vacuum expectation values, leading to the possibility of axion domain walls and attendant cosmological problems. (Similarly, our discrete anomalous symmetry may be combined not only with a $\alpha = \pi/2$ but also with $\alpha = \pi/2 + k \cdot 2\pi/6$ to give non-anomalous symmetries.) This problem is well known in axion physics [18]. At an abstract level it is most simply addressed by postulating that eq. (6.3) is not quite a symmetry of the full theory, but rather changes the phase θ of QCD instantons by $\theta \rightarrow \theta - 2\alpha$. This change leaves the basic PQ mechanism intact – any value of θ can be compensated by a suitable α – but removes the troublesome degeneracy. (The extra non-anomalous symmetry at $\alpha = \pi$ is related by a hypercharge transformation to an unbroken symmetry.)

The postulated additional change in θ with α could arise from the effect of very heavy quarks [19], as in the Kim scheme. There are also other ways of relieving the unwanted degeneracy.

7. Conclusions, and an additional comment

The spontaneous breaking of a discrete symmetry need not always be a cosmological disaster. One important discrete symmetry, used to ensure the absence of flavor-changing neutral currents in the two-doublet extension of the standard model, is safe by virtue of its being anomalous. Because of this its breakdown gives domain walls which are not quite stable, and leads to an interesting but not manifestly wrong cosmological scenario.

The nature of anomalous symmetry breaking is drastically modified by incorporation of the PQ mechanism. Depending on details of the PQ assignments, the domain walls may never form at all, or on the contrary may form and be stable.

We have also mentioned the possibility that a spontaneously violated discrete symmetry may secretly be embedded in a broken continuous gauge symmetry. (The meaning of this “secrecy” of course is that the gauge symmetry is generically badly broken, but is more mildly broken at the discrete point.) The breakdown of the

continuous symmetry leads to string formation. Unless these strings get inflated away, the domain walls formed subsequently are bounded by strings and decay rapidly. The situation is similar if the discrete symmetry is embedded in a continuous *global* symmetry. Then a network of *global* strings forms due to the breakdown of the continuous symmetry, and the domain walls formed subsequently are bounded by these strings and evolve without unacceptable consequences.

However, if inflation occurs in between the two phase transitions, then the strings are removed, and the secret continuous symmetry cannot prevent a cosmological catastrophe. This conclusion holds for either a continuous gauge symmetry or a continuous global symmetry. It may sound surprising at first that a global symmetry, in this context, is no more felicitous than a gauge symmetry. After all, a theory with a spontaneously broken continuous global symmetry has no mass gap, and one might expect the domain walls to decay by emitting Goldstone bosons. But this does not happen. The domain walls are classically stable, and decay only by nucleation of loops of (global) string, as described in sect. 3. The rate for this process is too small to save the cosmological scenario from disaster.

S.T. and F.W. would like to thank R.L. Davis for useful conversations.

References

- [1] Ya.B. Zel'dovich, I.Yu. Kobzarev and L.B. Okun, JETP 40 (1975) 1
- [2] S.L. Glashow and S. Weinberg, Phys. Rev. D15 (1977) 1958
- [3] R.D. Peccei and H.R. Quinn, Phys. Rev. Lett. 38 (1977) 1440; Phys. Rev. D16 (1977) 1791
- [4] S. Weinberg, Phys. Rev. Lett. 40 (1978) 223;
F. Wilczek, Phys. Rev. Lett. 40 (1978) 279
- [5] L.M. Krauss and F. Wilczek, Phys. Rev. Lett. 62 (1989) 1221;
J. Preskill and L.M. Krauss, Nucl. Phys. B341 (1990) 50
- [6] T.W.B. Kibble, G. Lazarides and Q. Shafi, Phys. Lett. B113 (1982) 237; Phys. Rev. D26 (1982) 435
- [7] A. Vilenkin and A.E. Everett, Phys. Rev. Lett. 48 (1982) 1867
- [8] J. Preskill, *in* Architecture of the fundamental interactions at short distances, ed. P. Ramond and R. Stora (North-Holland, Amsterdam, 1987) p. 235
- [9] D. Gross, R. Pisarski and L. Yaffe, Rev. Mod. Phys. 53 (1981) 43
- [10] T.W.B. Kibble, J. Phys. A9 (1976) 1387
- [11] A. Vilenkin, Phys. Rep. 121 (1985) 263
- [12] D. Stauffer, Phys. Rep. 54 (1979) 1
- [13] A.E. Everett, Phys. Rev. D10 (1974) 316
- [14] E. Witten, Phys. Rev. D30 (1984) 272
- [15] K.S. Thorne, Caltech preprint GRP-234 (1990)
- [16] J.E. Kim, Phys. Rev. Lett. 43 (1979) 103
- [17] M. Dine, W. Fischler and M. Srednicki, Phys. Lett. B104 (1981) 199;
A.P. Zhinitskii, Sov. J. Nucl. Phys. 31 (1980) 260
- [18] P. Sikivie, Phys. Rev. Lett. 48 (1982) 1156
- [19] H. Georgi and M.B. Wise, Phys. Lett. B116 (1982) 123
- [20] L. Ibáñez and G. Ross, Should discrete symmetries be anomaly free? CERN preprint TH 6000/91 (1991)