Discrete Gauge Symmetry in Continuum Theories

Lawrence M. Krauss[(a)] and Frank Wilczek[(b)]
Institute for Theoretical Physics, University of California at Santa Barbara, Santa Barbara, California 93106
(Received 16 December 1988)

We point out that local symmetries can masquerade as discrete global symmetries to an observer equipped with only low-energy probes. The existence of the underlying local gauge invariance can, however, result in observable Aharonov-Bohm-type effects. Black holes can therefore carry discrete gauge charges—a form of nonclassical "hair." Neither black-hole evaporation, wormholes, nor anything else can violate discrete gauge symmetries. In supersymmetric unified theories such discrete symmetries can forbid proton-decay amplitudes that might otherwise be catastrophic.

PACS numbers: 11.30.−j, 13.30.Ce, 97.60.Lf

Although it is a common and fruitful practice to consider local gauge invariance under discrete groups in lattice theories, the implications of such invariance in the continuum have not been widely discussed. (They have been invoked in one class of solutions to the axion domain-wall problem.[1,2])

At first sight the notion of local discrete symmetry in the continuum appears rather silly. Indeed, the most important dynamical consequence of a continuous local symmetry is the existence of a new field, the gauge field. This field is introduced in order to formulate covariant derivatives. Covariant derivatives are, of course, necessary so that invariant interactions involving gradients may be formed; such interactions in turn are necessary in order that charged fields may propagate. In the case of a discrete symmetry there is no similar need to introduce a gauge potential, because the ordinary derivative already transforms simply.

To make the discussion more concrete, let us consider a specific realization of the general idea of discrete local symmetry, where we produce a local Z_p symmetry. Consider a U(1) gauge theory containing two scalar fields η and ξ carrying charge pe and e, respectively. Suppose that η undergoes a condensation at some very high mass scale M, while ξ does not condense and produces quanta of relatively small mass. Then the effective low-energy theory will simply be the theory of the single complex scalar field ξ. This theory will be invariant under the transformation.

$$\xi \rightarrow e^{2\pi i/p}\xi \qquad (1)$$

as a consequence of the original gauge invariance. The only implication of the original gauge symmetry for the low-energy effective theory is the absence of interaction terms forbidden by Eq. (1). And this implication does not distinguish between local and global symmetry.

Nevertheless, there is a fundamental difference between local and global symmetries, whether continuous or discrete. It is that *global symmetry is a statement that the laws of physics take the same form when expressed in terms of various distinct variables, while local symmetry is a statement that the variables used in a physical theory are redundant.* In language that may be more familiar, this redundancy is often stated as the fact that in a gauge theory, only gauge-invariant quantities are physically meaningful.

From this point of view, it is clear that no processes, not even such exotic ones as black-hole evaporation or wormhole tunneling, can violate a gauge symmetry. There are two striking theoretical consequences of this observation:

(i) It has been argued recently that wormhole tunneling induces all local interactions consistent with *continuous* gauge symmetries.[3] (The restriction to continuous local symmetries is not always made explicitly, but has been tacitly assumed in the conclusions drawn.[4]) The theory of wormholes is presently in no fit state to supply quantitative estimates of the magnitude of the induced interactions. Still, something can be said. Plausibly, nonrenormalizable interactions induced by wormholes are suppressed by inverse powers of the Planck mass—or the wormhole scale, if this is different—but there is no evident small parameter suppressing renormalizable interactions. Taken at face value, this feature is a considerable embarrassment. For example, in models with low-energy supersymmetry, there are numerous renormalizable interactions which violate baryon number, and are capable of causing proton decay at a rapid rate.[5] Traditionally, such interactions have been argued away by invoking R parity or discrete flavor symmetries.[5] If wormholes made it impossible to maintain such symmetries, they would therefore create a great difficulty in reconciling the interesting possibility of low-energy supersymmetry with the stability of matter.[4] As another example, it is an attractive idea that the structure of the quark mass matrix is largely dictated by discrete symmetries.[6] This idea also appears to be endangered by wormholes.

In either case, promoting the relevant discrete symmetries to local symmetries would permit us to ensure that they are maintained, independent of the vicissitudes of wormhole dynamics.

Consider proton decay for example. It is straightforward, along the lines of Eq. (1), to introduce a local U(1) symmetry that, broken down to a discrete subgroup, prohibits worrisome dimension-4 trilinear quark superfield operators. Consider for example the U(1) symmetry associated with "fiveness," $=5(B-L)-4Y$,[7] which arises naturally in many supersymmetric unified models,[8] and which we know can be gauged without anomaly problems. By introducing scalar superfields with fiveness ± 4, we can easily break this symmetry down to fiveness (mod4) and still preserve supersymmetry. Left-handed quark and lepton chiral superfields are multiplied by i under the remaining discrete symmetry. Therefore, dimension-4 baryon-number–violating interactions (and also neutrino Majorana masses) will be forbidden.

(ii) It is commonly argued that charges associated with continuous local symmetries are the only meaningful characteristics of black holes (no-hair theorem[9]). We will now argue that discrete local symmetries supply others. For example, we will argue that the charge defined as an integer mod$p$, associated with the existence of a local Z_p symmetry described in Eq. (1), is an observable of black holes.

Once we widen our horizons to consider processes occurring at energies of order M, of course the underlying gauge degrees of freedom, if they exist, can be excited. Thus, by doing such experiments, we could learn that our low-energy discrete symmetry has secretly been a gauge symmetry. Since the Higgs field η in our initial example can only screen charges which are a multiple of pe, we could then infer the existence of a conserved charge modpe. Indeed, this charge is associated with the observable

$$\exp(2\pi i Q/pe) = \exp\left[2\pi i \int \mathbf{E}\cdot d\mathbf{s}/pe\right], \qquad (2)$$

where the equality expresses Gauss's law. In the Higgs phase, Q itself is ill-defined but the operator in Eq. (2) is well defined.[10,11] Now, since the right-hand side is expressed in terms of a surface integral, it cannot be affected by the gravitational collapse of what lies inside into a black hole; thus it is an intrinsic feature of the asymptotic black-hole state. We believe that a formal proof along these lines could be constructed by *defining* the black-hole state as a quantum state using disorder operators, as was done for soliton states by Frölich and Marchetti.[12]

A direct physical manifestation is also possible. The theory supports stable strings threaded by magnetic flux $2\pi/pe$. The scattering of ξ quanta, with charge e, from such strings is dominated at low energies by the Aharonov-Bohm effect.[13] The magnitude and form of this cross section is uniquely determined by the product of charge and flux, mod2π, and thus allows one, in principle, to make a precise observational determination of the Z_p-valued charge alluded to above. (Of course, postulating the existence of such strings takes us outside the framework of the effective low-energy theory as usually understood, so the existence of this effect does not really contradict the statement made above, that the low-energy effective theory does not distinguish local from global discrete symmetries.)

A simple thought experiment based on the Aharonov-Bohm scattering process allows us to demonstrate that black holes have discrete gauge hair. Let us imagine that we have a ξ quantum falling into a black hole, and let us scatter a string of very low energy and momentum from this composite object. The scattering cross section, which involves behaviors at large times and distances should not depend on the precise instant at which the particle crosses the event horizon—a rather fuzzy notion, in any case. And yet this cross section does depend critically on the Z_p charge. We must conclude that this Z_p charge does not depend on whether the particle has crossed the event horizon, and, in particular, that it retains its meaning (and induces the same Aharonov-Bohm phases) for the asymptotic, "pure" black hole.

The necessity that our local gauge symmetry *strictly* forbids couplings which violate the Z_p symmetry, argued on purely logical grounds above, also follows from physical considerations similar to those in the previous paragraph. The Aharonov-Bohm cross section would cease to be well defined if the discrete Z_p charge were not conserved.

Since the mass scale M, and the mass of the associated flux strings, can be made arbitrarily large, it is logical to ask whether we could take the limit and deal with the local Z_p symmetry in the continuum directly. What would it mean to do this? As we have emphasized, implementing the local symmetry does not require modifying the Lagrangian. Therefore, the Feynman rules remain the same. So what is the difference? Operationally, the difference between the local and global versions of this discrete symmetry amounts to a different prescription for carrying out functional integrals, which shows up only for large fluctuations of the fields. In the local symmetry version, one must consider that the ξ field takes values not in the complex plane C, but rather in the quotient of this space by the symmetry transformation Eq. (1). Now, tasteful quantum field theorists have traditionally been leery of the use of fields living in exotic spaces, at least in the context of four-dimensional theories, because of the difficulty of carrying out the usual renormalization program within them. (In two dimensions the situation is essentially different: Since scalar fields have zero mass dimension, arbitrary functions of them—including δ functions restricting them to exotic manifolds—still generate renormalizable theories.) However, since imposing discrete symmetries as envisaged here does not modify the local structure of field space, and the counterterms needed in renormalization theory are local in this sense, there is every reason to think that it does not lead to any significant difficulties. It seems appropriate, though, to mention two caveats. First, there are discrete symmetries—those associated with global anomalies—that cannot be consistently gauged. Identification of such anomalies is a difficult but well developed art,[14] into which we shall not enter here. Second, it is not quite true that the identifications we envisage in field space are locally trivial—the discrete transformations, in most interesting cases, have fixed points, leading to conical singularities. It is conceivable that these singularities lead to subtle problems that have not yet been discerned. The situation here is reminiscent of orbifold constructions in string theory,[15] which, in fact, our discrete local symmetries greatly resemble.

Indeed, it seems probable that many or all discrete symmetries that arise in effective theories derived from underlying string theories will be local, since such symmetries typically are just those few elements of the huge gauge symmetry groups [E(8)⊗E(8) and ten-dimensional general covariance] in the underlying theory that act trivially on all vacuum condensates. If so, then their validity will not be affected by the vicissitudes of wormhole dynamics. Also, black holes will have plenty of hair—perhaps just as much as any other elementary particle—so that the apparently sharp distinction between sufficiently heavy elementary particles and small black holes will fade away.

This research was supported in part by DOE and NSF.

[(a)]Permanent address: Center for Theoretical Physics, Sloane Lab, Yale University, New Haven, CT 06517. Also at Department of Astronomy, Yale University, New Haven, CT 06517 and Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138.

[(b)]Permanent address: School of Natural Sciences, Institute for Advanced Study, Princeton NJ 08540.

[1]G. Lazarides and Q. Shafi, Phys. Lett. **115B**, 21 (1982).

[2]A review containing many further references is J. Kim, Phys. Rep. **150**, 1 (1987).

[3]For a review of the already extensive literature on this subject, with references, see A. Strominger, University of California at Santa Barbara report, in Proceedings of the Theoretical Advanced Study Institute Summer School, 1988 (to be published).

[4]See especially G. Gilbert, Caltech Report No. CALT-68-1524 (to be published), which was the immediate stimulus to this work.

[5]S. Dimopoulos and H. Georgi, Nucl. Phys. **B193**, 150 (1981); E. Witten, Nucl. Phys. **B185**, 513 (1981); S. Dimopoulos, S. Raby, and F. Wilczek, Phys. Rev. D **24**, 1681 (1981); N. Sakai and T. Yanagida, Nucl. Phys. **B197**, 533 (1982); S. Weinberg, Phys. Rev. D **26**, 287 (1982); M. Dine *et al.*, Nucl. Phys. **B259**, 549 (1985).

[6]S. Weinberg, in *A Festschrift for I. I. Rabi,* edited by Lloyd Motz (New York Academy of Sciences, New York, 1977); F. Wilczek and A. Zee, Phys. Lett. **70B**, 418 (1977); H. Fritzsch, Phys. Lett. **73B**, 317–322 (1978).

[7]F. Wilczek and A. Zee, Phys. Lett. **88B**, 311 (1979).

[8]See Sakai and Yanagida, Ref. 5.

[9]E.g., see B. Carter, in *Black Holes, Les Houches 1972,* edited by C. DeWitt and B. S. DeWitt (Gordon and Breach, New York, 1973).

[10]G. Morchio and F. Stocchi, in *Fundamental Problems of Gauge Field Theory, Erice, 1985,* edited by G. Velo and A. S. Wightman (Plenum, New York, 1987).

[11]A. Goldhaber, R. Mackenzie, and F. Wilczek, Harvard Report No. HUTP-88/A044 (to be published).

[12]J. Fröhlich and P. A. Marchetti, Eidgenössische Technische Hochschule report, 1988 (to be published); also, Commun. Math. Phys. **116**, 127 (1988).

[13]Y. Aharonov and D. Böhm, Phys. Rev. **119**, 485 (1959); M. Alford and F. Wilczek, Phys. Rev. Lett. **62**, 1071 (1989).

[14]See *Current Algebra and Anomalies,* edited by S. B. Treiman, R. Jackiw, B. Zumino, and E. Witten (World Scientific, Singapore, 1987).

[15]L. Dixon, J. Harvey, C. Vafa, and E. Witten, Nucl. Phys. **B261**, 678 (1986).