

Anomalous Discrete Symmetries in Three Dimensions and Group Cohomology

Anton Kapustin

California Institute of Technology, Pasadena, California 91125, USA

Ryan Thorngren

University of California, Berkeley, California 94720, USA

(Received 11 March 2014; published 13 June 2014)

We study 't Hooft anomalies for a global discrete internal symmetry G . We construct examples of bosonic field theories in three dimensions with a nonvanishing 't Hooft anomaly for a discrete global symmetry. We also construct field theories in three dimensions with a global discrete internal symmetry $G_1 \times G_2$ such that gauging G_1 necessarily breaks G_2 and vice versa. This is analogous to the Adler-Bell-Jackiw axial anomaly in four dimensions and parity anomaly in three dimensions.

DOI: [10.1103/PhysRevLett.112.231602](https://doi.org/10.1103/PhysRevLett.112.231602)

PACS numbers: 11.10.Kk, 11.30.Er, 11.30.Qc

Introduction.—Since the discovery by Bell and Jackiw [1] and Adler [2] of the anomalous nonconservation of the axial current, anomalies have played an increasingly important role in particle physics. Recently anomalies found applications in condensed matter physics: they appear, implicitly or explicitly, in the classification of Symmetry Protected Topological (SPT) phases [3,4]. This viewpoint sheds a new light on anomalies and leads to some surprising conclusions. Motivated by these developments, we study anomalies of global discrete internal symmetries. In particular, we show that such anomalies can afflict bosonic field theories in odd space-time dimensions.

There are several different but related kinds of anomalies. The original ABJ discovery [1,2] was that a classical symmetry can be violated on the quantum level. We will call this phenomenon an ABJ anomaly. Anomalies can also affect gauge symmetries; gauge theories which suffer from such anomalies are inconsistent on the quantum level. Finally, it might happen that a global symmetry is consistent on the quantum level, but cannot be promoted to a gauge symmetry because the resulting gauge theory would be anomalous. In such a case one says that a global symmetry has an 't Hooft anomaly [5]. 't Hooft argued that 't Hooft anomalies of continuous symmetries are preserved under RG flow and thus constrain possible RG trajectories.

The source of all these anomalies is chirality: either the theory itself is chiral or the global symmetry acts in a chiral way. Since chiral matter exists only in even space-time dimensions, it is often said that in odd space-time dimensions anomalies are absent. However, in odd space-time dimensions there is another source of chirality, namely Chern-Simons couplings. This suggests that there may be anomalies whose existence depends on Chern-Simons interactions. It is this mechanism that causes the parity anomaly of three-dimensional gauge theories with fermions [6]. In these theories, maintaining gauge-invariance may

require adding a Chern-Simons interaction which breaks parity. This is an ABJ anomaly for a global space-time symmetry. We will show that a similar mechanism can lead to anomalies for global discrete internal symmetries in bosonic theories.

't Hooft anomalies and group cohomology.—ABJ and 't Hooft anomalies are closely related, and it is instructive to address the latter first. 't Hooft anomaly for a global symmetry group G is an obstruction for gauging G . If G is a connected Lie group, the form of the 't Hooft anomaly is tightly constrained by the Wess-Zumino consistency conditions [7]. They imply that in d space-time dimensions, possible anomalies for G are classified by Chern-Simons actions in dimension $d + 1$ [8]. An intuitive reason for this is as follows. On a $d + 1$ -manifold with a boundary, the Chern-Simons actions is gauge-invariant only up to boundary terms. To cancel the boundary terms, one has to couple the bulk theory to a d -dimensional boundary theory with an 't Hooft anomaly for G . This mechanism for canceling 't Hooft anomalies is called anomaly inflow. Conversely, if one assumes that an 't Hooft anomaly in d dimensions can be canceled by an anomaly inflow from $d + 1$ dimensions, then anomalies in d dimensions must be classified by possible Chern-Simons actions in $d + 1$ dimensions.

If G is not connected, Wess-Zumino consistency conditions are not as constraining. In the extreme case of a finite symmetry group G , they become vacuous. As a substitute, let us assume that the anomaly can be canceled by inflow from $d + 1$ dimensions. Then 't Hooft anomalies in d dimensions should be classified by topological actions in $d + 1$ dimensions [4]. For general G , such actions are classified by elements of the Abelian group $H^{d+2}(BG, \mathbb{Z})$ [9]. Here BG is the classifying space for G bundles [10]. Typically, BG is an infinite-dimensional space defined up to homotopy equivalence. For example, for $G = U(1)$ one can take $BG = \mathbb{C}P^\infty$. For finite G there is a nice explicit construction of BG [11].

The classification of topological actions in terms of cohomology of BG assumes that the action depends only on the gauge field. For fermionic systems the action may also depend on the spin structure, and then more complicated actions exist [4,9]. In this note we focus on bosonic systems and their anomalies.

From now on G will be a finite symmetry group. For finite G we have an isomorphism $H^{d+2}(BG, \mathbb{Z}) \simeq H^{d+1}(BG, U(1))$ [12]. Note that the same group classifies bosonic SPT phases with global symmetry G in $d+1$ dimensions [3]. The reason for this is as follows. Group cohomology classification of SPT phases relies on gauging G and integrating out everything except the gauge field for G . This gives an effective topological action for the G connection, and such actions for finite G are classified by $H^{d+1}(BG, U(1))$ [9]. Consequently, the boundary of an SPT phase classified by a cohomology class $\omega \in H^{d+1}(BG, U(1))$ must either break G spontaneously or carry a field theory with a global symmetry G which has an 't Hooft anomaly ω [3,4].

The anomaly inflow assumption does not always hold: there exist theories whose 't Hooft anomalies cannot be canceled by anomaly inflow and thus are not related to SPT phases [13]. Since our main goal here is to produce new examples of anomalous theories, we need not delve into this issue.

't Hooft anomalies in three dimensions.—In $d=3$ the relevant cohomology group is $H^4(BG, U(1))$. This cohomology group is nonvanishing for $G = \mathbb{Z}_n \times \mathbb{Z}_n$: $H^4(B(\mathbb{Z}_n \times \mathbb{Z}_n), U(1)) = \mathbb{Z}_n \times \mathbb{Z}_n$. (In contrast, $H^4(B\mathbb{Z}_n, U(1)) = 0$ for all n [12].) Thus there should exist three-dimensional field theories with a global symmetry group $\mathbb{Z}_n \times \mathbb{Z}_n$ which cannot be gauged.

To produce an example of such a theory we use the insight of Ref. [14], which argued that on a gapped boundary of an SPT phase the global symmetry G must either be broken or realized projectively. That is, relations between generators of G hold only modulo elements of a gauge group N . Then the symmetry of the theory is not a product $G \times N$, but an *extension* of G by N . By definition, an extension of G by N is any group \hat{G} which has N as a normal subgroup and such that $\hat{G}/N = G$. N has to be a normal subgroup, because conjugation by any element of \hat{G} must map a gauge symmetry to a gauge symmetry. We will be interested in the special case when N is Abelian and G acts trivially on the gauge fields. Then every element of \hat{G} commutes with every element of N , and one says that \hat{G} is a central extension of G by N . Central extensions are classified by the cohomology group $H^2(BG, N)$ [12]. Note that even if both G and N are Abelian, \hat{G} may be non-Abelian.

Suppose that the full symmetry group \hat{G} is a central extension of G by an Abelian gauge group N . In general, the fact that G acts projectively need not lead to 't Hooft anomalies. Let $N = \mathbb{Z}_2$, $\hat{G} = \mathbb{Z}_4$ and $G = \mathbb{Z}_2$. Here G is

generated by a single element g satisfying $g^2 = n$, where n is the generator of N . This is an example of a nontrivial extension of \mathbb{Z}_2 by \mathbb{Z}_2 . It is easy to construct models where the global \mathbb{Z}_2 acts as above but nevertheless can be gauged. For example, consider a $U(1)$ gauge theory coupled to a pair of scalars which transform as a doublet of $U(2)$. The global symmetry is $U(2)/U(1) = SU(2)/\mathbb{Z}_2$. Suppose further that the model has a singlet scalar of charge 2 which condenses at a high energy scale and spontaneously breaks $U(1)$ to \mathbb{Z}_2 . The global $SU(2)/\mathbb{Z}_2$ contains a \mathbb{Z}_2 subgroup generated by a transformation

$$g: \phi_{\pm} \mapsto \pm i\phi_{\pm}, \quad g^2 = -1. \quad (1)$$

Clearly, there is no obstruction for gauging the whole $U(2)$, which includes the finite subgroup generated by g .

On the other hand, a projective action of G opens a possibility for a 't Hooft anomaly. To construct an example with a 't Hooft anomaly, consider $U(1)$ Chern-Simons theory at level $2n$ coupled to some scalar fields. The precise matter content is not important, but it must be such that the action is invariant under a $\mathbb{Z}_n \times \mathbb{Z}_n$ symmetry, and the relations between generators hold up to elements of the gauge group $U(1)$. For example, we may consider n scalar fields of charge 1 on which the generators x and y of $\mathbb{Z}_n \times \mathbb{Z}_n$ act as $n \times n$ ‘‘clock’’ and ‘‘shift’’ matrices U and V satisfying

$$U^n = 1, \quad V^n = 1, \quad VU = e^{2\pi i/n}UV. \quad (2)$$

This central extension of $\mathbb{Z}_n \times \mathbb{Z}_n$ is the generator of $H^2(BG, U(1)) = \mathbb{Z}_n$. More generally, we may postulate that x and y act as U^q and V^q for some integer q , corresponding to the extension parameter $q \in H^2(BG, U(1))$.

Consider now gauging the $\mathbb{Z}_n \times \mathbb{Z}_n$ symmetry by coupling the theory to a $\mathbb{Z}_n \times \mathbb{Z}_n$ gauge field. We will regard $\mathbb{Z}_n \times \mathbb{Z}_n$ as a subgroup of $U(1) \times U(1)$ and represent the $\mathbb{Z}_n \times \mathbb{Z}_n$ gauge field by a pair of $U(1)$ gauge fields A_1, A_2 . These fields are flat and their holonomy must take values in n th roots of unity. Therefore we must have $nA_i = d\phi_i$, $i = 1, 2$, where $\phi_{1,2}$ are 2π -periodic scalars. These scalars have a transparent physical meaning: they can be regarded as phases of complex Higgs fields whose vacuum expectation values break $U(1) \times U(1)$ down to $\mathbb{Z}_n \times \mathbb{Z}_n$. That is, under the gauge transformations $A_i \mapsto A_i + df_i$, the scalars transform as

$$\phi_i \mapsto \phi_i + nf_i. \quad (3)$$

Note that while A_1 and A_2 are flat, they are not necessarily globally well-defined one-forms. Indeed, the above constraints allow the transition functions to be arbitrary n th roots of unity. Accordingly, the first Chern class for A_i can be nontrivial (but it becomes trivial when multiplied by n ; i.e., it is an n -torsion cohomology class).

The most general topological action describing the coupling of the Chern-Simons gauge field a to A_1, A_2 is

$$S_3 = \frac{1}{2\pi} \int \left(nada + p_1 A_1 da + p_2 A_2 da + \frac{p_3 n^2}{2\pi} a A_1 A_2 \right) + F(A_1, A_2), \quad (4)$$

where the local functional F depends only on $A_{1,2}$. Here we use the differential form notation for gauge fields. The coefficients p_1 and p_2 are $\mathbb{Z}_n \times \mathbb{Z}_n$ charges of a Dirac monopole for a and therefore must be integral. The coefficient p_3 must also be integer for $\exp(iS_3)$ to be invariant under the $U(1)$ gauge transformation $a \mapsto a + df$, where f is an arbitrary 2π -periodic scalar. It remains to determine the transformation properties of a under $\mathbb{Z}_n \times \mathbb{Z}_n$ gauge symmetry. In our approach, the symmetry is enhanced to $U(1) \times U(1)$, but this is compensated by the presence of two periodic scalars ϕ_i which transform as in (3). Requiring the action be invariant up to terms which are independent of a forces the following transformation laws:

$$a \mapsto a - \frac{p_3}{4\pi} f_1 d\phi_2, \quad a \mapsto a + \frac{p_3}{4\pi} f_2 d\phi_1. \quad (5)$$

These are well defined provided p_3 is even. These unusual transformation laws reflect that $\mathbb{Z}_n \times \mathbb{Z}_n$ is realized projectively. Indeed, the commutator of the above transformations is a $U(1)$ gauge transformation for a with the parameter $f = p_3 n f_1 f_2 / 4\pi$. Thus the generators x and y of $\mathbb{Z}_n \times \mathbb{Z}_n$ do not commute, and their commutator is a $U(1)$ gauge transformation of the form $\exp(if) = \exp(2\pi i p_3 / 2n)$. That is, the symmetry group \hat{G} is a central extension of $\mathbb{Z}_n \times \mathbb{Z}_n$ with the extension parameter $q = p_3 / 2$.

We can now compute how the action (4) transforms under $\mathbb{Z}_n \times \mathbb{Z}_n$ gauge transformations. For the first \mathbb{Z}_n we get

$$S_3 \mapsto S_3 + \frac{p_3 n}{8\pi^2} \int f_1 (p_1 A_2 dA_1 + p_2 A_2 dA_2) + F(A_1 + df_1, A_2) - F(A_1, A_2), \quad (6)$$

for the second \mathbb{Z}_n we get

$$S_3 \mapsto S_3 - \frac{p_3 n}{8\pi^2} \int f_2 (p_1 A_1 dA_1 + p_2 A_1 dA_2) + F(A_1, A_2 + df_2) - F(A_1, A_2). \quad (7)$$

It is easy to check that one cannot choose the local functional F so as to cancel both variations. This shows that the theory has an 't Hooft anomaly. On the other hand, one can cancel the anomaly by coupling the 3d theory to a 4d gauge theory with gauge group $\mathbb{Z}_n \times \mathbb{Z}_n$ and an action

$$S_4 = -\frac{np_3}{8\pi^2} \int_{M_4} A_1 A_2 (p_1 dA_1 + p_2 dA_2), \quad (8)$$

where M_4 is a four-dimensional manifold whose boundary is the 3d space-time. One can show that this continuum action describes Dijkgraaf-Witten theory for the gauge group $G = \mathbb{Z}_n \times \mathbb{Z}_n$. The corresponding element in $H^4(BG, U(1)) = \mathbb{Z}_n \times \mathbb{Z}_n$ is $(p_3 p_1 / 2, p_3 p_2 / 2)$. (An analogous continuum description of the two-dimensional Dijkgraaf-Witten theory has been studied in detail in [15]).

ABJ anomalies in three dimensions.—ABJ anomalies are closely related to 't Hooft anomalies. For example, one can interpret the famous axial anomaly as arising from an 't Hooft anomaly for $U(1)_{em} \times U(1)_A$. In general, suppose a global symmetry group has the product form $G = G_1 \times G_2$. Suppose also that G has a nontrivial 't Hooft anomaly, while both G_1 and G_2 have a trivial 't Hooft anomaly. Mathematically, this means that the cohomology class $\omega \in H^4(BG, U(1))$ becomes trivial when restricted to BG_1 or BG_2 (for example, because $H^4(BG_1, U(1)) = H^4(BG_2, U(1)) = 0$). Then G_1 or G_2 can be gauged, but not the whole G . What is the fate of G_2 in a theory with a gauged G_1 ? Either it is still a global symmetry or it is not. If it is a symmetry, then it must have a nontrivial 't Hooft anomaly. But if $H^4(BG_2, U(1))$ is trivial, this is impossible. Hence G_2 is not a symmetry; i.e., gauging G_1 necessarily breaks G_2 , regardless of the precise form of the action for the G_1 gauge field. Conversely, gauging G_2 necessarily breaks G_1 . We may regard this as a form of ABJ anomaly.

The example with $G = \mathbb{Z}_n \times \mathbb{Z}_n$ from the previous section has exactly the right structure, since $H^4(B\mathbb{Z}_n, U(1)) = 0$. To be concrete, we set $n = 2$, $p_1 = 1$, $p_2 = 0$, $p_3 = 2$. Thus we have a $U(1)$ Chern-Simons theory coupled to a pair of complex scalar fields of charge 1. Generators x and y of $\mathbb{Z}_2 \times \mathbb{Z}_2$ global symmetry act on scalar fields as Pauli matrices σ_3 and σ_1 . We may also have other fields on which $\mathbb{Z}_2 \times \mathbb{Z}_2$ acts trivially. In addition, $p_1 = 1$ implies that a Dirac monopole for the $U(1)$ gauge field is odd under the \mathbb{Z}_2 generated by x . There is no obstruction for gauging this \mathbb{Z}_2 . Let us now show that any consistent gauging breaks the second \mathbb{Z}_2 . Both the original $U(1)$ gauge symmetry and the first \mathbb{Z}_2 are subgroups of $U(1) \times U(1)$ which acts on the two scalar fields by diagonal matrices. As a preliminary step, let us gauge this $U(1) \times U(1)$ symmetry, and let b_1 and b_2 be the corresponding gauge fields. Then we spontaneously break the gauge group down to $U(1) \times \mathbb{Z}_2$ by condensing an additional scalar with $U(1) \times U(1)$ charge $(2, -2)$ and neutral under $\mathbb{Z}_2 \times \mathbb{Z}_2$. As a result of the gauge symmetry breaking, we are left with a $U(1)$ gauge field a and a \mathbb{Z}_2 gauge field A_1 (i.e., A_1 is a $U(1)$ gauge field such that $2A_1 = d\phi$ for some 2π -periodic scalar ϕ). In the low-energy theory the fields b_1 and b_2 can be parametrized as follows:

$$b_1 = a, \quad b_2 = a + A_1.$$

Our goal is to obtain a topological action for a and A_1 which has the form (4) with $A_2 = 0$ and the desired values of the parameters, i.e.,

$$S_3 = \frac{1}{2\pi} \int (2ada + A_1 da) + \dots,$$

where dots denote terms which do not depend on a . On the other hand, the most general Chern-Simons action for b_1 and b_2 has the form

$$\sum_{i,j} \frac{K_{ij}}{4\pi} \int b_i db_j,$$

where K_{ij} is a symmetric integral matrix. Expressing b_1 and b_2 in terms of a, A_1 , we find the following constraints on K_{ij} ,

$$K_{11} + K_{22} + 2K_{12} = 4, \quad 2K_{22} + 2K_{12} = 2,$$

which implies $K_{11} - K_{22} = 2$. Thus the Chern-Simons couplings of b_1 and b_2 must be unequal.

On the other hand, the second \mathbb{Z}_2 symmetry acts by σ_2 and exchanges the two scalar fields which couple to b_1 and b_2 , respectively. Thus it must also exchange b_1 and b_2 . But since K_{11} and K_{22} are unequal, the Chern-Simons couplings necessarily break the second \mathbb{Z}_2 symmetry. Since the breaking of $G_2 = \mathbb{Z}_2$ is due to topological terms in the action, it can be regarded as a quantum effect analogous to the ABJ anomaly. The analogy with the parity anomaly of three-dimensional gauge theories is even closer. There gauge-invariance forces one to choose the Chern-Simons level k to be half-integral, excluding the parity-invariant value $k = 0$ [6]. Similarly, in our example $U(1) \times G_1$ gauge invariance together with the requirement that the topological action for $U(1) \times G_1$ took the desired form forces us to break the remaining global $\mathbb{Z}_2 = G_2$.

Concluding remarks.—The 't Hooft anomalies for continuous symmetries are not affected by the RG flow and therefore lead to constraints on the IR behavior of theories [5]. This is true even if the symmetry in question is spontaneously broken, since breaking a continuous symmetry results in Goldstone bosons, and a 't Hooft anomaly gives rise to Wess-Zumino-Witten terms in their effective action [8].

For discrete symmetries the situation is different, since spontaneous breaking of discrete symmetries does not lead to massless particles. A phase with a discrete global symmetry G broken down to nothing is always possible and has trivial 't Hooft anomaly, regardless of the 't Hooft anomaly of the UV theory. But if some subgroup $G_0 \subset G$ remains unbroken, the same argument as for continuous symmetries shows that 't Hooft anomalies for G_0 must be the same in the UV and the IR.

Further, we saw that even massive QFTs can have nontrivial 't Hooft anomalies, if at long distances they reduce to sufficiently complicated TQFTs. Thus a non-vanishing 't Hooft anomaly in general does not rule out a gapped phase with an unbroken symmetry: it only rules out a trivial gapped phase with an unbroken symmetry.

Note that in [14] it was proposed that phases with this property (gapped phases with a symmetry G which, while unbroken, requires topological order) are precisely the gapped surface phases of bosonic SPTs in one dimension higher. But we saw above that such phases may have trivial 't Hooft anomalies. One such example is $G = \mathbb{Z}_2 \times \mathbb{Z}_2$, $N = \mathbb{Z}_2$, and a trivial topological action for the N -valued gauge field. In this case, G may act projectively on the scalars, in which case the \mathbb{Z}_2 gauge symmetry cannot be broken without breaking some of G . Topological \mathbb{Z}_2 order in such a theory is “protected” by the global discrete symmetry G . Nevertheless, such a theory has a trivial 't Hooft anomaly and therefore the corresponding SPT phase in four dimensions is trivial in the group cohomology classification.

We are grateful to T. Senthil, E. Witten, N. Seiberg, V. Oshikawa, and P. Etingof for discussions. The work of A. K. was supported in part by the DOE Grant No. DE-FG02-92ER40701.

Note added.—After this paper appeared on the electronic preprint archive, we learned that related results have been obtained by Cho *et al.* [16].

-
- [1] J. S. Bell and R. Jackiw, *Nuovo Cimento A* **60**, 47 (1969).
 - [2] S. L. Adler, *Phys. Rev.* **177**, 2426 (1969).
 - [3] X. Chen, Z.-C. Gu, Z.-X. Liu, and X.-G. Wen, *Phys. Rev. B* **87**, 155114 (2013); L.-Y. Hung and X.-G. Wen, *Phys. Rev. B* **87**, 165107 (2013).
 - [4] X.-G. Wen, *Phys. Rev. D* **88**, 045013 (2013).
 - [5] G. 't Hooft, NATO Adv. Study Inst. Ser. B Phys. **59**, 135 (1980).
 - [6] A. N. Redlich, *Phys. Rev. Lett.* **52**, 18 (1984).
 - [7] J. Wess and B. Zumino, *Phys. Lett.* **37B**, 95 (1971).
 - [8] S. Weinberg, *The Quantum Theory of Fields. Vol. 2: Modern Applications* (Cambridge University Press, Cambridge, England, 1996).
 - [9] R. Dijkgraaf and E. Witten, *Commun. Math. Phys.* **129**, 393 (1990).
 - [10] Cohomology of BG is also known as group cohomology of G . It should not be confused with the cohomology of G regarded as a topological space.
 - [11] J. Milnor, *Ann. Math.* **63**, 430 (1956).
 - [12] L. Evens, *The Cohomology of Groups* (Oxford University Press, Oxford, 1991).
 - [13] A. Kapustin and R. Thorngren, arXiv:1404.3230 [cond-mat.str-el].
 - [14] A. Vishwanath and T. Senthil, *Phys. Rev. X* **3**, 011016 (2013)
 - [15] A. Kapustin and N. Seiberg, *J. High Energy Phys.* **04** (2014) 001.
 - [16] G. Y. Cho, J. C. Teo, S. Ryu, arXiv:1403.2018 [cond-mat.str-el]