Noninvertible Peccei-Quinn Symmetry and the Massless Quark Solution to the Strong CP Problem

Clay Córdova,[1,2] Sungwoo Hong,[3] and Seth Koren[1,4]

[1]Enrico Fermi Institute, University of Chicago, Chicago, Illinois, USA
[2]Kadanoff Center for Theoretical Physics, University of Chicago, Chicago, Illinois, USA
[3]Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
[4]Department of Physics and Astronomy, University of Notre Dame, South Bend, Indiana, USA

(Received 4 April 2024; revised 11 March 2025; accepted 1 May 2025; published 10 July 2025)

We consider theories of gauged quark flavor and identify noninvertible Peccei-Quinn symmetries arising from fractional instantons when the resulting gauge group has nontrivial global structure. Such symmetries exist solely because the standard model has the same number of generations as colors, $N_g = N_c$, which leads to a massless down-type quark solution to the strong CP problem in an ultraviolet $SU(9)$ theory of quark color-flavor unification. We show how the Cabibbo-Kobayashi-Maskawa flavor structure and weak CP violation can be generated without upsetting our solution.

DOI: 10.1103/PhysRevX.15.031011 Subject Areas: Particles and Fields

I. INTRODUCTION

It has been 100 years since the birth of quantum physics, and it is still uncertain how long it will take to reach maturity. However, for a century, we have known the true mechanical laws that govern our Universe. What is the status of our understanding of the fundamental physical nature of the matter from which we are composed? The main conclusion is that the standard model of particle physics is a simple theory that represents the truly remarkable success of reductionism. It is astounding that we can understand the structure of the Universe down to distances of about 10^{-19} m, and the standard model description works exceedingly well.

The subleading story is that the structure of the standard model itself remains mysterious, and physicists have long been interested in pushing the program of reductionism even further: We wish to uncover a theory that explains why we exist at this point in the standard model's parameter space. The SM requires a couple dozen numbers that we must input, including a variety of angles and some very large ratios of scales. Beyond general qualitative considerations (cf. Dirac [1]), the structure of quantum field theory characterizes how difficult it is, in general, to have an ultraviolet theory that explains the variety of numbers needed (cf. 't Hooft [2]), and the SM poses certain challenges.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

The multiplicity of inputs arises mainly from the Yukawa sector, where three separate generations of matter means there are many masses and mixing angles. The Yukawa couplings themselves have the benefit that they are "technically natural," which suggests that the question of their sizes can be solved at small distances. The couplings y_u, y_d, and y_e (and y_ν with neutrino masses) are the sole breakings of the $U(3)^5$ flavor symmetries in three independent directions, so a spurion analysis tells us that, in the SM, they evolve with scale proportionally to themselves $\delta y_i \propto y_i$. This means we may hope to begin with (some of) these as exact symmetries and produce the needed sizes of yukawa couplings at some high scale, without this solution being disrupted by low-energy physics.

Indeed, in Ref. [3], we focused on the technically natural but ludicrously small Dirac neutrino Yukawas in the low-energy SM with right-handed neutrinos, where $y_{\nu_3}/y_\tau \lesssim 10^{-11}$, comparing the heaviest charged and neutral leptons. [4] With only the SM gauge group, the neutrino Yukawas are spurions for a familiar, invertible symmetry. However, an analysis of generalized symmetries in the lepton sector found that the neutrino Yukawas can be protected by a noninvertible symmetry in lepton flavor gauge theories like $U(1)_{L_\mu - L_\tau}$. This subtle interplay between the physics of lepton flavor monopoles and neutrino masses then guided us to a theory that produces these small numbers and is fully Dirac natural.

We wrote down an ultraviolet $SU(3)_H$ theory wherein all global symmetries are either good classical symmetries or explicitly broken by $\mathcal{O}(1)$ numbers like y_τ. Then, instanton effects of the ultraviolet theory break a classical symmetry and thus naturally produce small Dirac neutrino Yukawas from a Dirac natural theory $y_\nu \sim y_\tau \exp(-8\pi^2/g_H^2)$.

In understanding this model, we effectively began creating a roadmap of model building using noninvertible symmetries by showing how a technically natural, invertible spurion could be upgraded to a noninvertible spurion and then given a fully natural origin.

However, the standard model also contains parameters that do not have the protection of technical naturalness, and there is one such problem at each mass dimension 0, 2, and 4. Supersymmetry, which could have protected the cosmological constant (CC), is broken at a scale of at least $\Lambda_{\mathrm{SUSY}}/\Lambda_{CC} \gtrsim \mathrm{TeV}/\mathrm{meV} \sim 10^{15}$ larger than the energy scale of the CC, and it presents a major challenge to understanding the far infrared in our normal framework of Wilsonian effective field theory. Likewise, the Higgs mass $\mu^2 H^\dagger H$ has no protective symmetry in the SM, which indicates the danger of the electroweak hierarchy problem—first realized as the doublet-triplet splitting problem of grand unification but an important issue more generally in any UV field theory that introduces additional scales giving rise to $\delta\mu^2 \propto \Lambda^2$. Finally, the CP-violating phase in the strong sector, $\bar{\theta}$, is renormalized by the other source of quark CP violation, the Cabibbo-Kobayashi-Maskawa (CKM) phase δ_{CKM}. This lack of technical naturalness signals the strong CP problem.

One may naively wonder if generalized symmetries can be useful for more serious naturalness problems of the SM where we cannot rely on the low-energy theory being technically natural. In this paper, we show that this is indeed the case: As we will discuss below, the protective symmetries of the strong CP problem are quite subtle in the SM; we find that a noninvertible symmetry can shed light on a UV model that is fully Dirac natural, and the smallness of the strong CP angle is explained. Admittedly, strong CP is special, and the strong phase is closely related to the technically natural parameters of the Yukawa sector; thus, we do not wish to give the impression that solving the big, dimensionful hierarchy problems is close at hand. Yet it is intriguing that we can go further than one might expect, and we do not know where else generalized symmetries may lead us. For other applications of generalized global symmetries in particle physics, see, e.g., Refs. [3,5–31]; for recent reviews, see Refs. [32–37].

In this work, we examine the generalized symmetries of the quark sector and identify a noninvertible symmetry in the standard model when extended by a gauged horizontal symmetry of quarks that has nontrivial global structure. This process will require a more sophisticated generalized symmetry than our earlier work, where the noninvertible symmetry structure involved a $U(1)^{(1)}$ magnetic 1-form symmetry and could be located from familiar triangle anomaly computations. Here, the noninvertible symmetry will connect a discrete $\mathbb{Z}_3^{(1)}$ magnetic 1-form symmetry with a 0-form symmetry of quarks, and uncovering this structure will require us to examine our field theory on $S^2 \times S^2$ to find the relevant fractional instantons. The payoff will be discovering a generalized symmetry that exists specifically because the SM bears out $N_c = N_g$, with the same number of colors as generations.

After finding this noninvertible symmetry and identifying the spurions of the symmetry structure, our understanding of the breaking of generalized symmetries then provides us as infrared effective field theorists with additional model-building guidance. In particular, we learn that an ultraviolet theory that includes $\mathbb{Z}_3$ color-flavor monopoles necessarily breaks this noninvertible symmetry explicitly and thus generates nonvanishing values of the spurions from nonperturbative gauge theory effects. Having identified the down-type Yukawas as spurions for such a symmetry, this analysis tells us that noninvertible symmetry breaking in an ultraviolet theory containing these monopoles can potentially revive a "massless down-type quark" solution to the strong CP problem using small instantons. We are then pointed to a particularly simple theory of $SU(9)$ color-flavor unification that can solve strong CP in this manner without requiring any new fermions.

It is remarkable that such theories of nontrivial gauge-flavor unification are possible with the SM structure [38], and an intense study of their full phenomenologies is surely merited. We attempt to factorize the various physics effects here—for reasons we will discuss later, slightly richer structure is needed to find SM flavor, and we postpone a more detailed discussion to future work. Here, our main goal is to understand and explain the intriguing nonperturbative physics such models possess, as informed by the generalized symmetry analysis, and to show how such models can solve the strong CP problem. We hope our work motivates increased activity on the perturbative structure of these theories, including detailed flavor model building, as well as consideration of complementary signatures such as their early Universe cosmologies.

In the rest of this introduction, we briefly review the strong CP problem as well as the original massless up-quark solution. Then, in Sec. II, we provide a generalized symmetry analysis in the infrared theories of gauged quark flavor. In Sec. III, we introduce the ultraviolet model of $SU(9)$ color-flavor unification; we explain how it solves the strong CP problem in the UV and how this can be preserved while generating the CKM matrix. Various further directions are discussed in Sec. IV.

Some technical and pedagogical material is reserved for the appendixes. In Appendix A, we review the construction of fractional instantons. In Appendix B, we review the global structure of gauge groups. In Appendix C, we discuss one-loop corrections to 't Hooft vertices.

A. Strong *CP* problem

The strong CP angle is the field-redefinition-invariant CP-odd phase combining the color topological density with the quark Yukawa determinants,

$$\bar{\theta} = \arg\, e^{-i\theta} \det(y_u y_d), \tag{1.1}$$

where $i\theta/32\pi^2$ is the coefficient of $G\tilde{G}$ in the Lagrangian; if $\bar{\theta} = 0$; there is a field basis in which the quark Yukawas have real eigenvalues. Impressive searches for, and constraints on, the electric dipole moment of the neutron indeed give us the upper bound $\bar{\theta} \lesssim 10^{-10}$ [39,40]. If $\bar{\theta}$ were the only CP-violating parameter in the theory of quarks, then it would be technically natural, as spurious CP symmetry would imply that small $\bar{\theta}$ is stable under renormalization group (RG) flow $\delta\bar{\theta} \propto \bar{\theta}$.

However, there is more CP violation allowed in the quark Yukawas for $N_g \geq 3$ generations. The other CP-violating parameter is the "CKM phase" δ_{CKM}, which arises from the misalignment of the quark Yukawas with the weak interaction basis and can be invariantly parametrized by the other CP-odd field-redefinition-invariant combination of Yukawas, the Jarlskog invariant, [41]

$$\tilde{J} = \mathrm{Im}\, \det[y_u^\dagger y_u, y_d^\dagger y_d]. \tag{1.2}$$

The size of $\tilde{J}$ is often understood by going to the mass basis where the Yukawas are diagonal and quark mixing is in the CKM matrix, and then parametrizing the CKM matrix in terms of mixing angles and a single CP-violating phase, $\delta_{\mathrm{CKM}} \sim 1.14$ [43]. This parameter is responsible for the CP violation originally observed in decays of neutral kaons [44].

While the CKM phase obeys Dirac's naturalness principle in being an $\mathcal{O}(1)$ angle, $\bar{\theta}$ is not technically natural, and in the limit of small $\bar{\theta}$, its renormalization group evolution goes as

$$\delta\bar{\theta} \propto c\,\delta_{\mathrm{CKM}}. \tag{1.3}$$

Thus, it seems there is no protection of $\bar{\theta}$ against quantum mechanical destabilization. Now, it is numerically true that the coefficient c in the standard model is very small [45] as a result of the many approximate symmetries, which we will discuss further in the next section. Given only the SM content, the running from Eq. (1.3) does not result in an infrared $\bar{\theta}$ in excess of the constraints.

Still, the qualitative difference of $\bar{\theta}$ not being the only CP-odd spurion indicates that it can be a general concern in further UV theories. Indeed, ultraviolet field theories can easily allow operators that could quickly introduce additional CP-violating phases into the quark sector and destabilize $\bar{\theta}$ far more than δ_{CKM}. Since $\bar{\theta}$ is not the only CP-violating parameter, simply imposing a CP symmetry on the ultraviolet theory is not straightforwardly available.

This naturalness tension has motivated decades of work on how physics beyond the standard model might square these two facts. One approach is to forge ahead by imposing a discrete symmetry in the ultraviolet (e.g., CP *à la* Nelson-Barr [46–48] or parity [49–52]), and by clever model-building explain how δ_{CKM} is generated while $\bar{\theta}$ is not. These models have many interesting phenomenologies and observational signatures. They also require many new colored fermions (sometimes introduced in very particular structures), and they are often in danger of being destabilized at loop order (see, e.g., Refs. [16,53–58] for various concerns).

An alternative approach takes advantage of θ being quite special as a CP-odd spurion, in that it can also be a spurion for anomalous chiral symmetries. The relative benefit of such strategies is that they do not need to impose a discrete spacetime symmetry at any energy scale. Rather, the structure of the theory is such that one particular CP-violating parameter, the strong CP angle, is naturally set to zero regardless of the CP violation present elsewhere in the theory, and we now turn to explaining this structure.

Other approaches to strong CP can be found in, for example, Refs. [59–72], among many others. Useful reviews include Refs. [73–76].

B. Peccei-Quinn for strong *CP*

The sort of solution to the strong CP problem that we will pursue engages directly with the fact that the theta angle can have an intricate spurionic structure in a theory of colored fermions, such that CP is not the only symmetry at play.

Consider an $SU(3)_C$ gauge theory that includes some θ topological angle, [77]

$$S \supset i\frac{\theta}{8\pi^2}\int_{\mathcal{M}} \mathrm{Tr}(G \wedge G), \tag{1.4}$$

with G the strong gauge field strength. Obviously, θ is an odd spurion for P or CP because the Levi-Civita symbol is a pseudotensor, but this is not necessarily the only spurionic charge for θ. We consider adding to our theory some colored fermions generically denoted as $\{q_i\}$. Then, for any global $U(1)_j$ symmetry rotation by an angle α_j acting on the fermions as

$$U(1)_j\colon\ q_i \to q_i e^{i g_{ij}\alpha_j}, \tag{1.5}$$

for some integer charges g_{ij}, if it has an Adler-Bell-Jackiw (ABJ) [78,79] anomaly,

$$\mathcal{A}_{SU(3)_C^2\times U(1)_j} \equiv \mathcal{A}_j = \sum I_i g_{ij}, \tag{1.6}$$

where I_i is the Dirac index of fermion q_i, then such a rotation of the quarks does not leave the theory invariant. Rather, it effects a change in the partition function as

$$Z \to Z \exp\left(i\mathcal{A}_j \frac{\alpha_j}{8\pi^2}\int_{\mathcal{M}} \mathrm{Tr}(G \wedge G)\right). \tag{1.7}$$

This transformation effectively modifies the θ angle, or in other words, θ has become a spurion transforming nonlinearly under $U(1)_j$: $\theta \to \theta + \mathcal{A}_j\alpha_j$. In the quark sector, such anomalous symmetries are often known as Peccei-Quinn (PQ) symmetries [80,81]. If $U(1)_j$ is a good symmetry, then the partition function remains invariant under an α_j transformation, so the physics does not depend on the value of θ.

Now, the effect of the anomaly is not just to give θ a spurionic charge, and the true structure here is a bit subtle. If there is an anomaly, $\mathcal{A}_j \neq 0$, then $U(1)_j$ is never truly a good, invertible symmetry—the ABJ anomaly from non-Abelian instantons always implies explicit breaking of order proportional to $\exp(-8\pi^2/g^2)$. However, in an asymptotically free gauge theory, $g^2 \to 0$ at high energies, and the instanton violation decouples quickly; thus, it makes sense to talk about $U(1)_j$ as a good symmetry in this limit. Then, as one evolves to low energies, explicit violation of $U(1)_j$ appears from the dynamical growth of the gauge coupling. Solutions to the strong CP problem involving PQ symmetries make intrinsic use of this low-energy violation of what began as a good UV symmetry.

The axion approach is to introduce a complex scalar ϕ that is charged under $U(1)_{PQ}$ and spontaneously breaks it. When ϕ obtains a vev, its angular degree of freedom a is a Goldstone for $U(1)_{PQ}$, so it shifts nonlinearly under PQ symmetry transformations like θ. In pure QCD, it is known quite generally that the gauge effects produce a potential $V(\theta)$ minimized at $\theta = 0$ [82]. Now, including an axion modifies the gluon action schematically as $i\int(\theta + a)G\tilde{G}$; thus, QCD instantons generate a PQ-violating but CP-preserving potential $V(a+\theta)$ that is minimized at $a = -\theta$. While, in pure QCD, $V(\theta)$ only tells us about the relative vacuum energy of different theories with different θ parameter values, a is now a dynamical degree of freedom.

If $U(1)_{PQ}$ began as a good symmetry, then this instanton effect generates the only potential for the axion. Thus, a can cosmologically relax to this solution to "screen" whatever CP-violating θ angle was present and allow the observed strong CP violation to vanish. Unfortunately, the minimal version of this model, the Weinberg-Wilczek axion [83,84], which directly couples ϕ to the SM quarks and spontaneously breaks one of the SM PQ symmetries, has long since been ruled out. Axion models are revived by adding new, vectorlike, colored fermions with their own $U(1)_{PQ}$ symmetries that ϕ can spontaneously break. These "invisible axion" models have been a subject of increased interest in recent years, so there is no need to mention that they have rich, fascinating phenomenological signatures. However, using axions is not the only way to take advantage of a good $U(1)_{PQ}$ to solve the strong CP problem.

The idea of the massless up quark solution was to instead posit that a SM $U(1)_{PQ}$ is *not* spontaneously broken. Since the up quark has the lightest mass in the infrared, one can imagine an ultraviolet symmetry acting on the up quark in such a way as to forbid a mass term,

$$U(1)_{\rm PQ}\colon\ \bar{u}_1 \to \bar{u}_1 e^{i\alpha} \Rightarrow \det y_u = 0. \tag{1.8}$$

Such a good symmetry implies that strong CP violation vanishes, but there is a slight subtlety in how we talk about it, given the standard definition of $\bar{\theta}$ in Eq. (1.1). Let us define the complex parameter $M \in \mathbb{C}$, which is the field-redefinition-invariant combination of the quark Yukawa eigenvalues,

$$M \equiv e^{-i\theta}\det(y_u y_d), \tag{1.9}$$

where the definition of the strong CP phase above is simply $\bar{\theta} \equiv \arg M$. The confusion is that this definition fails when $|M| \to 0$, such as when $U(1)_{PQ}$ is imposed. Of course, when the magnitude vanishes, the phase is undefined, which is sometimes discussed as $\bar{\theta}$ becoming unphysical in such a scenario. However, this language merely encodes an artifact of using polar coordinates to parametrize M. Alternatively, we could work in Cartesian coordinates; then, we have the CP-odd spurion

$$\mathrm{CP}\colon\ \mathrm{Im}(M) \to -\mathrm{Im}(M), \tag{1.10}$$

which manifestly behaves smoothly as $|M| \to 0$. Clearly, CP is preserved when $\mathrm{Im}(M) = 0$, and $\bar{\theta} = 0$ corresponds to $M \in \mathbb{R}_+$. When we have a good $U(1)_{PQ}$ symmetry, $|M| \to 0$; then, we could say $\bar{\theta}$ is unphysical, or we can just say that the CP-odd mass parameter vanishes, $\mathrm{Im}(M) = 0$.

Now that we have discussed this subtlety, let us return to the massless up-quark solution, having begun in the UV with such a $U(1)_{\rm PQ}$ symmetry. As we have long known from current algebra and hadron masses, the obvious issue is that the up-quark mass does not vanish in the infrared—$U(1)_{PQ}$ has been broken. However, this is sensible since it is an anomalous symmetry. Just as QCD instantons violating $U(1)_{PQ}$ provide a potential for the axion that localizes $\bar{\theta}$ to the CP-conserving value, they could also potentially accord a good UV PQ symmetry with the observed quark masses. Indeed, the contributions of instantons to the masses of quarks automatically preserve the form of the CKM matrix in which $\bar{\theta} = 0$ [85–88]. In other words, the instanton effects will violate $|M| = 0$ but only along the real axis. Whatever phases appear in the Yukawas, we continue to have $\mathrm{Im}(M) = 0$ as a basis-independent statement. Then, the idea is that m_u might be zero in the UV, corresponding to a high-quality $U(1)_{PQ}$, which is broken solely by QCD instantons that provide the observed $m_u > 0$.

After the massless up-quark solution was proposed in the mid-1980s, the observational status of this idea was held in limbo for some decades because the analytic calculation of instanton effects in QCD is not under theoretical control [89,90]. Eventually, numerical computations of QCD on the lattice became powerful enough to resolve whether a vanishing up-quark mass could fit data. Alas, the standard model does not bear out the massless up-quark solution [91,92]. In other words, one must begin at energies $\Lambda \gg \Lambda_{QCD}$ with some nontrivial Yukawa for the up quark already present in order to fit the far-infrared observables.

In recent years, it has been realized that this solution may be revived in UV completions of the SM in which $SU(3)_C$ emerges from a larger gauge group in which it is non-trivially embedded [93,94] (see also earlier work on the possible relevance of small color instanton effects, e.g., Refs. [95–97]). In some cases, instantons from the UV scale of the breaking $H \to SU(3)_C$ can provide a dominant contribution to the mass of the up quark (or the axion potential [98,99]). This possibility has been studied in flavor deconstruction [93] and in composite Higgs models [100].

Here, we identify extensions of the SM with gauged quark flavor symmetries in which θ becomes a spurion for a noninvertible Peccei-Quinn symmetry. This understanding of the generalized symmetry structure of the standard model reveals a minimal realization of a small instanton approach to the massless quark solution in the context of color-flavor unification. That such nontrivial structures are even possible with the SM chiral matter content is very suggestive. This approach will have various benefits, as we will see below, including the possibility to separate the scales of the instanton effects at a UV symmetry-breaking scale and the flavor-breaking effects at the scale $H \to SU(3)_C$. By starting with more of the SM's approximate global symmetries as gauge symmetries, we begin with an extremely simple theory, and we must understand how the SM structure is generated. However, unification has tremendous reductionist appeal, and the grand challenge in this model will be no more and no less than understanding a fully predictive theory of the quark Yukawas. In this work, we factorize issues and show how this new strong CP solution will work whether or not we obtain the mass hierarchies and mixing angles exactly, which we will leave to a future pursuit.

II. MASSLESS QUARKS FROM NONINVERTIBLE PECCEI-QUINN SYMMETRIES

We seek to generate Yukawa couplings by breaking noninvertible chiral symmetries in certain extensions of the SM that result from gauging an (approximate) global symmetry of the SM fields. In the lepton sector, this analysis leads to a model of neutrino masses [3]. In the quark sector, such chiral symmetries might be termed "noninvertible Peccei-Quinn symmetries," and we investigate them by examining anomalies of SM fields. The material we present and our setup are aimed toward the strong CP problem, but our analysis of noninvertible symmetries may potentially be of broader interest. The result of the noninvertible symmetry analysis will be to learn that certain parameters of a theory protected by noninvertible symmetry are necessarily generated by nonperturbative gauge theory effects in a UV embedding of the SM fields that introduces quark color-flavor monopoles.

The analysis of this section is brief and self-contained. A deeper analysis of noninvertible chiral symmetry can be found in Refs. [101,102]. We also provide a review of fractional instantons associated with nontrivial global structure in Appendix A, as well as a discussion on the global structure of the SM gauge group in Appendix B [103].

A. Noninvertible chiral symmetry

To start, let us briefly recall some aspects of noninvertible chiral symmetry [101,102] that will appear below. These symmetries often arise in chiral gauge theories. The simplest example concerns a current J_μ for a chiral symmetry that is violated by an ABJ anomaly:

$$\partial^\mu J_\mu = \frac{k}{32\pi^2} F_{\alpha\beta} F_{\gamma\delta} \epsilon^{\alpha\beta\gamma\delta}, \tag{2.1}$$

where, in the above, $k \in \mathbb{Z}$ is an integral anomaly coefficient and $F_{\alpha\beta}$ is the field strength of an Abelian gauge field. The current J_μ is no longer conserved, but as is well known, the absence of Abelian instantons implies that, at the level of the S-matrix (or, relatedly, local operator correlation functions), the chiral symmetry selection rules are still enforced. Noninvertible symmetry provides a way for us to understand this phenomenon nonperturbatively and to understand the unique features of such chiral symmetries. The key idea is to recognize that the right-hand side of Eq. (2.1) is a composite operator, which is itself built from symmetry currents. Indeed, the Bianchi identity implies that the field strength operator is closed, or more explicitly,

$$\partial^\mu (\varepsilon_{\alpha\beta\mu\nu} F^{\alpha\beta}) = 0. \tag{2.2}$$

This finding signals the presence of a higher symmetry: the magnetic 1-form symmetry of an Abelian gauge theory [105]. The charged objects under this symmetry are 't Hooft lines, which physically represent the worldlines of probe magnetic monopoles.

The appearance of a higher-form symmetry current in the anomaly equation (2.1) enables us to construct the operator (symmetry defect) that performs finite chiral symmetry transformations on Hilbert space. Let Σ denote the spatial slice (fixed time locus) where we wish to transform the fields by a chiral rotation by a finite angle $2\pi/kN$ for integral N. If the symmetry were invertible ($k = 0$), then we would simply construct a symmetry defect operator for an arbitrary phase α as

$$U_\alpha[\Sigma] = e^{i\alpha \int_\Sigma J}, \tag{2.3}$$

where we recognize $Q[\Sigma] \equiv \int_\Sigma J$ as the Noether charge. With the ABJ anomaly, it fails to be topological, and a new ingredient is needed. In addition to the exponentiated integrated current, along Σ, we must also include new topological degrees of freedom to cancel the ABJ anomaly and define a consistent conserved charge. These topological degrees of freedom are charged under a 1-form symmetry and hence can couple to the bulk through the magnetic 1-form symmetry current discussed above. In more detail, the Lagrangian on Σ for the additional topological degrees of freedom involves a new dynamical $U(1)$ gauge field C_μ and takes the form of a Chern-Simons action:

$$\mathcal{L}_\Sigma = \frac{iN}{4k}\int_\Sigma d^3x C_\mu \partial_\nu C_\sigma \epsilon^{\mu\nu\sigma} + \frac{i}{2\pi}\int_\Sigma d^3x C_\mu \partial_\nu A_\sigma \epsilon^{\mu\nu\sigma}, \quad (2.4)$$

where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. More generally, this construction can be carried out for any finite chiral transformation by a rational angle, resulting in a conserved charge (topological operator) that can implement the desired chiral transformation. The consequence of coupling the additional topological degrees of freedom on the worldvolume means, in modern terminology, that the symmetry has become noninvertible. In other words, it is no longer represented by unitary operators acting on Hilbert space.

The noninvertible nature of the chiral symmetry transformation has important technical and physical consequences. To illustrate these consequences, let us denote by $\mathcal{D}_{kN}(\Sigma)$ the operator on Σ described above, which implements a finite chiral transformation, and let $\bar{\mathcal{D}}_{kN}(\Sigma)$ denote the transformation by the opposite angle. When composed, these transformations do not equal unity but instead leave behind a condensate of magnetic 1-form symmetry operators [101,102,106–108]:

$$\mathcal{D}_{kN}(\Sigma) \times \bar{\mathcal{D}}_{kN}(\Sigma) \sim \sum_{\text{two-cycles } S\subset\Sigma} \exp\left(\frac{i}{16\pi N}\int_S F_{\alpha\beta} dS_{\mu\nu}\epsilon^{\alpha\beta\mu\nu}\right). \quad (2.5)$$

One term on the right-hand side, corresponding to a trivial cycle S, is the unit operator, which is the naive result of the multiplication. It is the only relevant term when acting on local operators. The remaining terms on the right-hand side are visible only when acting on states carrying magnetic charge or, equivalently, on 't Hooft line operators.

The interplay of magnetic charge and the chiral symmetry transformation is the inevitable result of the anomaly equation (2.1). Indeed, while there are no Abelian instantons in spacetimes with trivial topology, in a richer geometry such as those produced effectively by magnetic charges, Abelian instantons are possible and the anomaly can be saturated. More formally, it is natural to topologically view spacetime $\mathbb{R}^4$ with an infrared regulator as a four-sphere S^4. The absence of Abelian instantons is then a consequence of the fact that there are no topologically nontrivial two-cycles and hence no locus where a magnetic flux can be nontrivial. By contrast, including a 't Hooft line effectively modifies the spacetime topology: The two-sphere linking the worldline of the charge is now a nontrivial two-cycle, and, in general, in such configurations, Abelian instantons exist. Thus, the noninvertible symmetry analysis above gives a formal way to understand selection rules that arise from the absence of instantons on S^4. Note that this analysis also reveals the importance of the fact that the right-hand side of Eq. (2.1) is composed of Abelian field strengths. In general, for non-Abelian field strengths, instanton configurations exist already on S^4, and hence there are no resulting noninvertible chiral symmetries.

A crucial physical consequence of the fusion algebra in Eq. (2.5) is that it reveals a channel for noninvertible symmetry breaking. Consider a model with dynamical magnetic monopoles. Below the scale of their mass, the worldline of the monopoles appears in the infrared effective Abelian gauge theory as a 't Hooft line. Above the higgsing scale, the monopoles reveal themselves as fluctuating modes, and the magnetic 1-form symmetry is broken [109]. In other words, at this scale, Eq. (2.2) no longer holds. However, since the magnetic 1-form charges appear in the algebra of the chiral symmetries (2.5), they in turn must also be broken by the presence of magnetic charges. At a practical level, this means that loops of magnetic monopoles break the chiral symmetry. Since magnetic monopoles are solitons, the effects of such loops are generally nonperturbative in gauge theory couplings. To estimate their size, we parametrize their mass as

$$m_{\text{mon}} \sim \frac{v}{g}, \quad (2.6)$$

where v is the Higgs vev and g the gauge coupling. A monopole loop exists for a characteristic proper time $\delta\tau$ that is inversely proportional to the effective cutoff set by the W-boson mass and hence $\delta\tau \sim 1/(gv)$. The corrections from monopole loops therefore have a characteristic size:

$$\delta\mathcal{L} \sim \exp(-S_{\text{mon}}) \sim \exp(-m_{\text{mon}}\delta\tau) \sim \exp(-\#/g^2). \quad (2.7)$$

Thus, when the chiral symmetry violation is mediated by 1-form symmetry breaking effects of magnetic monopoles, the above estimate yields the expected size of the leading corrections.

We also note that Eq. (2.7) is precisely the characteristic size of non-Abelian instanton effects. Indeed, loops of magnetic monopoles—and, more generally, dyons—can be viewed as describing instanton corrections to the action [11], which is natural since a dyon has nonvanishing $E \cdot B$ and hence can saturate the topological term in the action. The breaking of a noninvertible chiral symmetry by magnetic monopole loops can therefore alternatively be understood as the direct breaking of chiral symmetry in the non-Abelian gauge theory by instanton effects. Conversely, the lens of noninvertible symmetry provides a key tool to understanding when instantons will yield the leading contribution to a given physical process and can thus guide us toward interesting UV models. We will utilize this model-building perspective below.

Thus far, we have seen how noninvertible chiral symmetry provides a language for dealing with Abelian ABJ anomalies. In our application below, we will instead require a more sophisticated version of this construction that goes beyond the physics of Abelian gauge theories. Specifically, we will employ noninvertible chiral symmetries that exist due to the presence of discrete magnetic 1-form symmetries in non-Abelian gauge theories.

In general, discrete magnetic 1-form symmetries occur in non-Abelian gauge theories where the gauge group has a nontrivial fundamental group, which we often refer to as a nontrivial global structure. For instance, the gauge group $SU(N)$ does not have any discrete magnetic 1-form symmetry, while the gauge group $SU(N)/\mathbb{Z}_N$ has magnetic symmetry $\mathbb{Z}_N$. Physically, the gauge group $SU(N)/\mathbb{Z}_N$ has a richer spectrum of magnetic monopoles (realized as nondynamical 't Hooft lines), which carry conserved $\mathbb{Z}_N$ quantum numbers. Crucial for our purposes, discrete magnetic 1-form symmetry also signals the presence of fractional instantons. The hallmark of such field configurations is that the instanton number is no longer an integer, but instead, its fractional part can be expressed as a suitable square of a discrete magnetic flux [see, e.g., Eq. (2.21)]. As above, this feature implies that, in a topologically trivial setup—without two-cycles or, equivalently, 't Hooft lines—the minimal allowed instanton number cannot be saturated. Indeed, on $\mathbb{R}^4$ (or, more precisely, the IR regulated S^4), non-Abelian instanton numbers are always integral.

In the context of ABJ anomalies and chiral symmetries, fractional instantons therefore have an effect similar to the Abelian instantons discussed above. In particular, there can again be chiral symmetries where the minimum allowed anomaly coefficient cannot be saturated on S^4. In this situation, the (typically discrete) chiral symmetry is then, in fact, noninvertible. The symmetry defect operator supports a fractional Hall state, which now couples to the bulk non-Abelian gauge fields through the discrete magnetic flux. (For more explicit formulas, see Ref. [102].) Thus, discrete noninvertible chiral symmetries depend, in detail, on the precise global form of the gauge group and can, in turn, point toward specific UV completions where these symmetries are then violated by new loops of dynamical monopoles or, equivalently, new small instanton effects. We exploit these mechanisms in our models below.

B. Approximate symmetries of the standard model

We now apply these considerations in the SM and beyond. We recall the basic structure of the SM fermions as reviewed in Table I. Throughout, we use conventions where hypercharge is integrally normalized.

The SM includes the Yukawa interactions coupling the fermions to the Higgs field H:

$$\mathcal{L} \supset y_u^{ij}\tilde{H}Q_i\bar{u}_j + y_d^{ij}HQ_i\bar{d}_j + y_e^{ij}HL_i\bar{e}_j, \tag{2.8}$$

we $\tilde{H} \equiv i\sigma_2 H^\star$. The observed Yukawa matrices y (equivalently, the fermion masses and flavor changing processes) explicitly break all of the non-Abelian continuous global symmetries, as they provide different masses for the generations. We first consider models that have vanishing down-type Yukawas $y_d \to 0$ and later aim to regenerate these couplings through symmetry-breaking effects. Our analysis will reveal that the down-type quark Yukawa y_d can be protected by certain kinds of noninvertible symmetries. We note that the same analysis with $y_u \leftrightarrow y_d$ instead would give the same conclusions, but our analysis will ultimately be the right choice for our purposes. At the classical level, nonvanishing, general y_u leaves the following Abelian flavor symmetries unbroken:

$$\prod_i^3 U(1)_{\tilde{B}_i} \times U(1)_{\bar{d}_i}, \tag{2.9}$$

where $\tilde{B}_i$ and the conventional baryon number B_i are defined as

$$\tilde{B}_i = Q_i - \bar{u}_i, \qquad B_i = \tilde{B}_i - \bar{d}_i = Q_i - \bar{u}_i - \bar{d}_i. \tag{2.10}$$

In fact, the full $U(3)_{\bar{d}}$ is thus far a full symmetry, but it is the analysis of the $U(1)$ factors that is pertinent below. Here and elsewhere, we will abuse notation in reusing the symbols for the SM fields to also refer to the charges of $U(1)$ symmetries that act on those species.

In Table II, we list the Adler-Bell-Jackiw [78,79] anomaly coefficients of these global symmetries with the SM gauge group and additional gauge groups appearing in

TABLE I. Representations of the standard model Weyl fermions under the classical gauge and global symmetries. We normalize each $U(1)$ so the least-charged particle has unit charge. We also list the charges of the right-handed neutrino N and the Higgs boson H.

	Q_i	$\bar{u}_i$	$\bar{d}_i$	L_i	$\bar{e}_i$	N_i	H
$SU(3)_C$	**3**	$\bar{\mathbf{3}}$	$\bar{\mathbf{3}}$	⋯	⋯	⋯	⋯
$SU(2)_L$	**2**	⋯	⋯	**2**	⋯	⋯	**2**
$U(1)_Y$	+1	−4	+2	−3	+6	⋯	−3
$U(1)_B$	+1	−1	−1	⋯	⋯	⋯	⋯
$U(1)_L$	⋯	⋯	⋯	+1	−1	−1	⋯

TABLE II. ABJ anomalies of chiral symmetries of the quark sector with SM gauge groups and a gauged $U(1)_H \equiv U(1)_{B_1+B_2-2B_3}$. We also show the anomaly coefficients in the fractional instanton background, denoted as [CH] (for "color-H" admixture), appearing in the $SU(3)_C \times U(1)_H/\mathbb{Z}_3$ extension.

	$U(1)_{\bar{B}_1}$	$U(1)_{\bar{B}_2}$	$U(1)_{\bar{B}_3}$	$U(1)_{\bar{d}_1}$	$U(1)_{\bar{d}_2}$	$U(1)_{\bar{d}_3}$
$SU(3)_C^2$	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$
$SU(2)_L^2$	N_c	N_c	N_c	0	0	0
$U(1)_Y^2$	$-14N_c$	$-14N_c$	$-14N_c$	$4N_c$	$4N_c$	$4N_c$
$U(1)_H^2$	N_c	N_c	$4N_c$	N_c	N_c	$4N_c$
$U(1)_Y U(1)_H$	$-2N_c$	$-2N_c$	$4N_c$	$-2N_c$	$-2N_c$	$4N_c$
[CH]	1	1	2	1	1	2

extensions discussed below. One immediate lesson we can extract from Table II is that there exists no noninvertible symmetry within the quark sector with gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y/\Gamma$ with $\Gamma = 1$. This is because each of these $U(1)$ global symmetries and any linear combinations of them are dominantly broken by non-Abelian instantons, or they remain good invertible symmetries, e.g., $\bar{d}_1 - \bar{d}_2$. Indeed, as reviewed in Sec. II A above, noninvertible chiral symmetry arises from a classical $U(1)$ dominantly broken either by an Abelian instanton or a fractional instanton of non-Abelian gauge theory. As we show in Appendix B 2, the absence of noninvertible symmetries of the quark sector with up Yukawas turned on remains the case even with nontrivial global structure $\Gamma \in \{\mathbb{Z}_6, \mathbb{Z}_3, \mathbb{Z}_2\}$.

So far, our analysis has revealed no surprising symmetries. With vanishing down-quark Yukawas, some PQ symmetries are restored, and hence strong *CP* is conserved. However, such a scenario, like the massless up-quark solution, is excluded since it fails to reproduce the observed IR physics of QCD. Instead, we consider two minimal extensions of the SM that enjoy noninvertible chiral symmetry acting on the quarks and suggest a natural UV completion. The new symmetries we find will protect the down-Yukawa couplings in such a way that UV symmetry breaking can generate the observed nonzero values while setting the strong *CP* phase $\bar{\theta}$ to zero.

Our extensions are based on gauging certain "horizontal" symmetries of the quark sector. In Sec. II C, we discuss a Z' extension by a $U(1)_H$ gauge group with

$$H = B_1 + B_2 - 2B_3, \tag{2.11}$$

and an extension by non-Abelian $SU(3)_H$ gauged horizontal flavor symmetry will be presented in Sec. II D. Each model below may be viewed as a separate candidate extension of the SM. Alternatively, it is also possible to think of them as two different phases of the same theory along a renormalization group flow. Starting from a UV

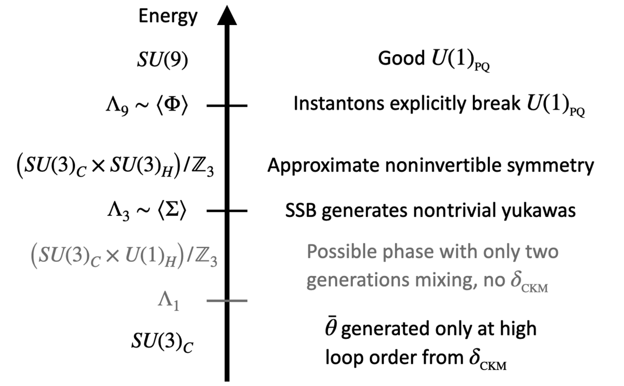

FIG. 1. For visual reference, the organization of phases governing the quarks in this work. In the model of Sec. III, we do not make use of the possible $U(1)_H$ phase, though there is also a noninvertible symmetry in this phase as discussed in Sec. II C.

theory [for instance, the $SU(9)$ model presented in Sec. III], one flows to the $SU(3)_H$ extension as an intermediate phase. Further flowing to the IR leads to the $U(1)_H$ extension. Finally, spontaneous breaking of $U(1)_H$ at yet a lower scale brings us to the SM (see Fig. 1). Ultimately, as discussed in Sec. III, to achieve the IR physics of the SM, we must break the new chiral symmetries under consideration. First, however, we present the approximate symmetries from the infrared as a key guide to our model.

C. $U(1)_H$ extension

One interesting possible Z' extension that is free of cubic and mixed anomalies (hence can be gauged with no additional matter) is the horizontal baryon number symmetry $U(1)_H$ given by Eq. (2.11). [110] Similar lepton family difference symmetries that are anomaly-free [e.g., $U(1)_{L_\mu - L_\tau}$] have received much attention as Z' models for reasons both formal and phenomenological. They are exact symmetries of the standard model with zero neutrino masses, but they are not exact symmetries of the real world. The similarity of the lepton and quark sectors in the SM (especially if neutrinos are Dirac type) suggests also investigating the quark family difference symmetries. Indeed, they are also approximate global symmetries of nature but are less well preserved in the infrared because the quark Yukawas are much larger. Further motivation for this particular choice of $U(1)_H$ will be given presently.

Before proceeding with a detailed analysis of the symmetries, we note a crucial discrete identification of the gauge and global symmetry groups. Specifically, with the matter content of the SM, the following groups act identically on all fields:

$$[\mathbb{Z}_3^H \subset U(1)_H] = [\mathbb{Z}_3^C \subset SU(3)_C] = [\mathbb{Z}_3^Y \subset U(1)_Y] = [\mathbb{Z}_3^B \subset U(1)_B]. \tag{2.12}$$

Thus, various diagonal $\mathbb{Z}_3$ subgroups act trivially on all fields, and we can modify the theory by quotienting the gauge group by any of these trivially acting subgroups. In our case, the relevant choice is between the gauge groups

$$SU(3)_C \times U(1)_H \quad \text{and} \quad (SU(3)_C \times U(1)_H)/\mathbb{Z}_3, \tag{2.13}$$

while other possible quotients do not play a role in our analysis. This choice can be clarified physically in the language of higher symmetries as follows. If we do not quotient by $\mathbb{Z}_3$, then the theory has an electric $\mathbb{Z}_3$ 1-form global symmetry whose charged objects are Wilson lines that cannot be screened by dynamical matter. By contrast, in the theory with the $\mathbb{Z}_3$ quotient, these representations are removed; correspondingly, Dirac quantization allows new magnetic monopoles whose 't Hooft lines are charged under a dual $\mathbb{Z}_3$ magnetic 1-form symmetry. Relatedly, the theory with the quotient admits fractional instantons. These effects will be crucial to our theory, and hence we focus on the case where the gauge group has the $\mathbb{Z}_3$ quotient [right-hand side of Eq. (2.13)]. We note that this possible gauge group global structure relies precisely on our choice of horizontal gauge group and exists due to the fact that $N_g = N_c$.

The ABJ anomaly coefficients of the global symmetries listed in Eq. (2.9) with $U(1)_H$ are given in Table II. In searching for noninvertible symmetries, we are interested in global $U(1)$ factors that are anomalous only due to the $U(1)_H$ effect [or at least where $U(1)_H$ provides the dominant anomalous breaking]. It is useful to notice that SM anomalies are flavor universal; therefore, $\tilde{B}_i - \tilde{B}_j$ and $\bar{d}_i - \bar{d}_j$, with any choice of i, $j = 1, 2, 3$, are SM-anomaly free. There are then four obvious candidates:

$$\begin{aligned} A_1 &= \bar{d}_3 - \bar{d}_1, \qquad & A_2 &= \bar{d}_3 - \bar{d}_2, \\ A_3 &= \tilde{B}_3 - \tilde{B}_1, \qquad & A_4 &= \tilde{B}_3 - \tilde{B}_2. \end{aligned} \tag{2.14}$$

However, we note that the combination

$$A_1 + A_2 - A_3 - A_4 = H \tag{2.15}$$

is gauged, so the generators in Eq. (2.14) above represent three possible flavor symmetries.

Consider first the effect of the familiar ABJ anomaly on these symmetries. From Table II, we have the anomaly coefficients

$$\begin{aligned} [U(1)_{A_{i=1,\dots,4}}][U(1)_H]^2 &= 3N_c, \\ [U(1)_{A_{i=1,\dots,4}}][U(1)_Y][U(1)_H] &= 6N_c, \end{aligned} \tag{2.16}$$

which breaks each $U(1)_{A_i}$ to a $\mathbb{Z}_{3N_c}$ invertible symmetry. Meanwhile, the rest of $U(1)_{A_i}$ forms an infinite noninvertible symmetry acting as rotations with a rational angle different from $\mathbb{Z}_{3N_c}$. In the case where the gauge group does not have a $\mathbb{Z}_3$ quotient, this concludes the analysis, and the down Yukawas are spurions of invertible symmetries. However, in the case with $(SU(3)_C \times U(1)_H)/\mathbb{Z}_3$, additional fractional instantons further modify these symmetries.

1. Symmetry breaking from fractional $(SU(3)_C \times U(1)_H)/\mathbb{Z}_3$ instantons

We now analyze the effects of fractional instantons on the symmetries. Focusing on $U(1)_H/\mathbb{Z}_3$, the first observation is that the magnetic flux is now fractionally quantized in units of one-third:

$$\int_\Sigma \frac{F_H}{2\pi} \in \frac{1}{3}\mathbb{Z}. \tag{2.17}$$

Those fluxes above which are integrally quantized are standard field configurations of $U(1)_H$, while those that are fractional are new configurations allowed by the quotient. More subtly, the $SU(3)_C/\mathbb{Z}_3$ also admits new flux configurations that are not allowed in $SU(3)_C$. These are labeled by a discrete analog of the magnetic flux, sometimes referred to as a second Stiefel-Whitney class:

$$\omega(A_C) \in H^2(M, \mathbb{Z}_3), \tag{2.18}$$

where M is the spacetime four-manifold and A_C is the color gauge field. Concretely, this means that ω is an object that may be integrated over any two-cycle Σ in spacetime, yielding an integer that is well defined modulo 3,

$$\int_\Sigma \omega(A_C) \in \mathbb{Z}_3, \tag{2.19}$$

and should be viewed as the non-Abelian analog of the fractional part of the magnetic flux in Eq. (2.17).

In our situation, the gauge group $(SU(3)_C \times U(1)_H)/\mathbb{Z}_3$ does not have independent quotients but rather one quotient that acts simultaneously on the two factors. This finding, in turn, implies that the fractional magnetic fluxes between the Abelian and non-Abelian factors are correlated, which we express as

$$\frac{F_H}{2\pi} = \frac{1}{3}\omega(A_C) - X. \tag{2.20}$$

Here, $X \in H^2(M, \mathbb{Z})$ can be viewed as a standard quantized flux, and each term can be thought of as a cohomology class (equality holds upon integration on any two-cycle).

The final technical tool we need is an analysis of the instanton number for gauge groups with discrete quotients. As reviewed in Appendix A, the quotient implies that the instanton number $\mathcal{N}_C$ of $SU(3)_C$ is no longer integral but is, in general, fractional, with the fractional part controlled by the discrete magnetic flux Eq. (2.19),

$$\mathcal{N}_C = \frac{1}{8\pi^2}\int_M \mathrm{Tr}(F_C \wedge F_C) = \frac{1}{3}\int_M \omega\wedge\omega \mod 1, \tag{2.21}$$

where the notation above means that the fractional parts of the left- and right-hand sides agree. Using Eq. (2.20), we can also express this case in terms of the flux F_H:

$$\begin{aligned}\mathcal{N}_C &= 3\int_M \left(\frac{F_H}{2\pi}+X\right)\wedge\left(\frac{F_H}{2\pi}+X\right) \mod 1\\ &= 6\left(\frac{1}{8\pi^2}\int_M F_H\wedge F_H\right) + 6\int_M \frac{F}{2\pi}\wedge X\\ &\quad + 3\int_M X\wedge X \mod 1\\ &= 6\left(\frac{1}{8\pi^2}\int_M F_H\wedge F_H\right) \mod 1\\ &= 6\mathcal{N}_H \mod 1,\end{aligned} \tag{2.22}$$

where, in the third line, we have used Eq. (2.20) to show that the contributions involving X do not modify the fractional part, and in the last line, we have introduced $\mathcal{N}_H$, which is the instanton density of $U(1)_H/\mathbb{Z}_3$:

$$\mathcal{N}_H = \frac{1}{8\pi^2}\int_M F_H\wedge F_H. \tag{2.23}$$

The final result of Eq. (2.22) shows that instantons of $(SU(3)_C\times U(1)_H)/\mathbb{Z}_3$ have a correlated fractional part; i.e., $\mathcal{N}_C$ can only be fractional if $\mathcal{N}_H$ is also fractional.

We can now use the above to compute the most refined anomaly coefficients in the presence of fractional instantons and discover the final fate of the symmetries in our problem. Using the notation $\psi_i = \{Q_i, \bar{u}_i, \bar{d}_i\}$ to denote a general charged fermion, the Dirac index for ψ_i in an instanton background is computed as

$$I_{\psi_i} = n_{\psi_i} T_{\psi_i}\mathcal{N}_C + \mathrm{dim}_{\psi_i} n_{\psi_i}(q^H_{\psi_i})^2\mathcal{N}_H, \tag{2.24}$$

where n_{ψ_i} is the multiplicity of ψ_i, including both flavor and $SU(2)_L$ gauge degrees of freedom. Here, T_{ψ_i} denotes the Dynkin index of ψ_i under $SU(3)_C$ (which is 1 for all ψ_i), dim_{ψ_i} is the dimension of the $SU(3)_C$ representation (which is 3 for all ψ_i), and finally, $q^H_{\psi_i}$ is the $U(1)_H$ charge. Importantly, even though the instanton numbers $\mathcal{N}_C$ and $\mathcal{N}_H$ are, in general, individually fractional, the indices above are always integral due to Eq. (2.22) and the fact that the matter content is consistent with the $\mathbb{Z}_3$ quotient. The anomaly coefficient for an Abelian flavor symmetry f in the presence of a general instanton background is then given by the formula

$$\begin{aligned}\mathcal{A}_f &= \sum_{\psi_i} q^f_{\psi_i} I_{\psi_i}\\ &= 3\mathcal{N}_H\sum_{\psi_i} q^f_{\psi_i} n_{\psi_i}(2+(q^H_{\psi_i})^2) + k\sum_{\psi_i} q^f_{\psi_i} n_{\psi_i},\end{aligned} \tag{2.25}$$

where we have used Eq. (2.22) as well as the details of our fermion spectra discussed above to simplify the index formula, and $k = \mathcal{N}_C - 6\mathcal{N}_H$ is an integer. The strongest constraints now come from choosing the most fractional instantons possible, which, from Eq. (2.17), are $\mathcal{N}_H = 1/9$ and $k = 0$ (so that $\mathcal{N}_C = 2/3$). Carrying out the sum then leads to the anomaly coefficients summarized in the final row of Table II.

From this analysis, one sees that, for each of the symmetries A_i defined in Eq. (2.14), we have

$$[U(1)_{A_{i=1,\ldots,4}}][\mathrm{CH}] = 1. \tag{2.26}$$

Therefore, the fractional instantons of $(SU(3)_C\times U(1)_H)/\mathbb{Z}_3$ completely turn each $U(1)_{A_i}$ into noninvertible symmetries, of which a discrete $\mathbb{Z}_3$ subgroup is particularly notable. Consider the following equality of charges modulo 3:

$$\tilde{B} + \bar{d} = H + A_1 + A_2, \quad (\mathrm{mod}\ 3). \tag{2.27}$$

As H is gauged, the diagonal flavor combination $A_1 + A_2$ generates a discrete flavor symmetry $\mathbb{Z}_3^{\tilde{B}+\bar{d}}$, which acts in a generation-independent way. According to our analysis above, $\mathbb{Z}_3^{\tilde{B}+\bar{d}}$ is a noninvertible symmetry. As we will see below, this discrete symmetry plays a key role in protecting quark masses.

More generally, our calculations of the anomaly coefficient also allow us to deduce the subgroup of invertible symmetries, i.e., those that do not participate in any anomalies involving H. A general charge J can be expressed in terms of integers ℓ_i as

$$J = \ell_1 A_1 + \ell_2 A_2 + \ell_3 A_3 + \ell_4 A_4. \tag{2.28}$$

From Eq. (2.26), the condition that J defines an invertible symmetry is then

$$\ell_1 + \ell_2 + \ell_3 + \ell_4 = 0. \tag{2.29}$$

Modding out by the H gauge redundancy leaves a rank-two invertible flavor symmetry. These symmetries are summarized in Table III.

2. Massless down quarks from noninvertible PQ symmetry

Having identified all symmetries of the theory, we now discuss the spurion structure of the down-quark Yukawa terms,

TABLE III. Symmetries of the standard model with no down Yukawas after gauging $U(1)_H$ with nontrivial global structure.

Gauged	$U(1)_{B_1+B_2-2B_3}$
Invertible	$U(1)_{\bar{B}_3-d_3-\bar{B}_1+d_1} \times U(1)_{\bar{B}_3-d_3-\bar{B}_1+d_2}$
Noninvertible	$U(1)_{\bar{d}_1+\bar{d}_2-2\bar{d}_3} \supset \mathbb{Z}_3^{\tilde{B}+\bar{d}}$

$$\mathcal{L}_{y_d} = y_d^{ij} H Q_i \bar{d}_j. \tag{2.30}$$

Of course, in our analysis above, we set $y_d^{ij} \to 0$. Thus, we can now see which symmetries must inevitably be broken to regenerate nonzero Yukawas.

First, we must enforce $U(1)_H$ gauge invariance:

$$U(1)_H \text{ allowed}: \; Q_1 d_1, Q_2 d_2, Q_3 d_3 \quad \text{(diagonal)} \tag{2.31}$$

$$Q_1 d_2, Q_2 d_1 \quad \text{(off-diagonal)}$$

$$U(1)_H \text{ forbidden}: \; Q_1 d_3, Q_3 d_1, Q_2 d_3, Q_3 d_2. \tag{2.32}$$

This result is a simple consequence of the disparity in $U(1)_H$ charge assignments between the first two generations and the third generation, which clearly implies there is no mixing of the third generation with the first or second.

Since $U(1)_{A_{1,2}}$ acts only on $\bar{d}$, it is easy to see that all $U(1)_H$-invariant components listed above are forbidden by noninvertible Peccei-Quinn symmetries. In particular, we note that all $U(1)_H$-allowed components of the Yukawa matrix are forbidden by the $\mathbb{Z}_3^{\tilde{B}+\bar{d}}$ noninvertible discrete symmetry. Thus, even if all other symmetries are broken, in the $U(1)_H$ phase with $\mathbb{Z}_3^{\tilde{B}+\bar{d}}$, the down-quark Yukawa matrix must vanish. Conversely, we may also view nonzero entries of the down Yukawa matrix as spurions of the noninvertible symmetry [111].

As discussed before and after Eq. (1.10), we see that when the noninvertible Peccei-Quinn symmetry is preserved, there are massless quarks and CP is a symmetry of the strong sector of the SM. For this reason, such models inform us about massless down-quark solutions to the strong CP problem. We note that this class of models—and, in particular, the tight interplay with noninvertible symmetry—only occurs for the choice of global form of the gauge group $(SU(3)_C \times U(1)_H)/\mathbb{Z}_3$, which in turn is only possible because $N_c = N_g$. Of course, in reality, down quarks are massive, implying that the noninvertible symmetry must be explicitly broken. The completion of the massless down-quark solution therefore requires a mechanism of noninvertible symmetry violation and the generation of observed quark masses, mixings, and the CKM phase, which we discuss in Sec. III.

D. $SU(3)_H$ extension

We now consider gauging the entire anomaly-free non-Abelian quark flavor symmetry $SU(3)_H$. We again make a nontrivial choice of quotient in the color-flavor gauge group:

TABLE IV. Symmetry and matter content of the $SU(3)_H$ extension.

	$SU(3)_C$	$SU(3)_H$	$U(1)_B$	$U(1)_{\bar{d}}$
$\mathbf{Q}$	3	3	+1	0
$\bar{\mathbf{u}}$	$\bar{3}$	$\bar{3}$	−1	0
$\bar{\mathbf{d}}$	$\bar{3}$	$\bar{3}$	−1	+1

$$\frac{SU(3)_C \times SU(3)_H}{\mathbb{Z}_3}. \tag{2.33}$$

The quarks transform in bifundamental representations summarized in Table IV. For instance, the quark doublet (now boldfaced) $\mathbf{Q}$ is both a 3 of $SU(3)_C$ unifying the color (red, green, and blue) quantum numbers as well as a 3 of $SU(3)_H$ unifying the flavor (e.g., up, charm, and top) quantum numbers as a gauge symmetry. That all colored SM matter obeys this pattern allows the nontrivial global structure chosen in Eq. (2.33).

Much of the symmetry analysis parallels that of the Abelian horizontal extension discussed above. We again consider the limit of vanishing down-type Yukawas and further restrict the up-type Yukawas to be family symmetric and hence compatible with $SU(3)_H$. The relevant classical symmetries are now flavor independent,

$$\frac{U(1)_B}{\mathbb{Z}_3} \times U(1)_{\bar{d}}, \tag{2.34}$$

which have charges listed in Table IV. Here, the $\mathbb{Z}_3$ quotient on $U(1)_B$ arises because that subgroup is gauged as noted in Eq. (2.12).

We first consider the anomalies that do not probe the global structure of the gauge group. The relevant anomaly coefficients are summarized in Table V. The net breaking effect is given by the greatest common divisor of all anomaly coefficients, which shows that the classical global symmetries are broken down to

TABLE V. ABJ anomalies of chiral symmetries of the quark sector with gauged $SU(3)_H$. We also show anomaly coefficients with fractional instantons, denoted as [CH] (for color-H admixture), allowed by the global structure $(SU(3)_C \times SU(3)_H)/\mathbb{Z}_3$.

	$U(1)_B$	$U(1)_{\bar{d}}$
$SU(3)_C^2$	0	N_g
$SU(2)_L^2$	$N_c N_g$	0
$U(1)_Y^2$	$-18 N_c N_g$	$4 N_c N_g$
$SU(3)_H^2$	0	N_c
[CH]	0	2

$$\frac{U(1)_B}{\mathbb{Z}_3} \times U(1)_{\bar{d}} \to \frac{\mathbb{Z}_9^B}{\mathbb{Z}_3} \times \mathbb{Z}_3^{\bar{d}} \cong \mathbb{Z}_3^B \times \mathbb{Z}_3^{\bar{d}}. \quad (2.35)$$

Without the quotient on the gauge group in Eq. (2.33), the above concludes the analysis of the symmetries. However, with the $\mathbb{Z}_3$ quotient, there are further instanton configurations to consider. Specifically, there are now gauge fields with nontrivial Stiefel-Whitney classes for both the color and horizontal gauge groups with

$$w_2(A_C) = w_2(A_H) \in H^2(M, \mathbb{Z}_3). \quad (2.36)$$

Instantons of both the color and horizontal gauge group can then have fractional instanton numbers valued in $\frac{1}{3}\mathbb{Z}$. However, because of Eq. (2.36), the fractional parts of the instanton numbers must be equal. More technically, the analog of Eq. (2.22) relating the fractional part of the instanton numbers is now

$$\mathcal{N}_C = \mathcal{N}_H, \quad \text{mod } 1. \quad (2.37)$$

To compute anomalies, we must evaluate the general sum of indices weighted by charges. For a general flavor symmetry f, the anomaly coefficient $\mathcal{A}_f$ is

$$\mathcal{A}_f = \sum_{\psi_i} q^f_{\psi_i} I_{\psi_i} = 3(\mathcal{N}_C + \mathcal{N}_H)(2q^f_Q + q^f_{\bar{d}} + q^f_{\bar{u}}). \quad (2.38)$$

Thus, we see that, in the minimal fractional instanton, for which $\mathcal{N}_C = \mathcal{N}_H = 1/3$, we have

$$\mathcal{A}_B = 0, \qquad \mathcal{A}_{\bar{d}} = 2. \quad (2.39)$$

Therefore, fractional instantons leave the baryon symmetry B untouched but turn $\mathbb{Z}_3^{\bar{d}}$ into a noninvertible symmetry.

As in our analysis in Sec. II C 2, we now arrive at a model where the down-type Yukawa coupling must vanish due to the presence of noninvertible chiral symmetry. Indeed, $SU(3)_H$ gauge invariance means that the Yukawa is reduced to a single number y_d:

$$\mathcal{L}_{y_d} = y_d \delta^{i\bar{j}} H Q_i \bar{d}_{\bar{j}}. \quad (2.40)$$

The down quarks transform under the noninvertible $\mathbb{Z}_3^{\bar{d}}$; hence, as long as this is a good symmetry, the down-type quarks are massless.

Just as in the $U(1)_H$ extension, we see that when the noninvertible $\mathbb{Z}_3^{\bar{d}}$ Peccei-Quinn symmetry is preserved, there are massless quarks, and CP is a symmetry of the strong sector of the SM. Moreover, this class of models exists only for the choice of global form of the gauge group $(SU(3)_C \times SU(3)_H)/\mathbb{Z}_3$, which in turn is only possible because $N_c = N_g$. We now turn to a UV completion of these models, which can break these symmetries and generate physical quark masses, mixings, and the CKM phase.

III. NONINVERTIBLE SYMMETRY BREAKING FROM COLOR-FLAVOR UNIFICATION

The noninvertible chiral symmetries of the infrared theories above point us to an ultraviolet theory where small instantons dynamically break these symmetries. The minimal choice is $SU(9)$ color-flavor unification, where the three colors and three generations of each quark field are intermingled in the fundamental of $SU(9)$. We will describe this theory in Sec. III A and show how its $U(1)_{PQ}$ symmetry protects down-quark masses and ensures that strong CP violation vanishes in the UV. In Sec. III B, we show that, at the scale Λ_9 where $SU(9) \to (SU(3)_C \times SU(3)_H)/\mathbb{Z}_3$, the instantons dynamically generate a flavor-symmetric y_d while keeping $\bar{\theta} = 0$ (further details on the 't Hooft vertices are given in Appendix C). In Sec. III C, we generate nontrivial flavor structure and weak CP violation at the scale Λ_3, where $(SU(3)_C \times SU(3)_H)/\mathbb{Z}_3 \to SU(3)_C$, while ensuring $\bar{\theta}$ continues to vanish, and in Sec. III D, we discuss determining Λ_9/Λ_3 from the running of the strong gauge coupling.

In the spirit of the massless up-quark solution to the strong CP problem—which does not work in the SM where the instanton effects are not large enough—our ultraviolet color-flavor unified $SU(9)$ theory contains additional instantons that can have larger effects in generating quark masses. An additional physics benefit is the possibility to separate the scale of instanton effects Λ_9 where $SU(9) \to SU(3)^2/\mathbb{Z}_3$ and the scale of flavor Λ_3 where $SU(3)_H$ is broken. Of course, this unification results in breaking of the noninvertible chiral symmetries of quarks discussed in detail in Sec. II, which make up the basic framework for this massless quark solution.

However, while our embedding in a flavor-unified gauge theory simplifies the UV description, on the contrary, it also makes the UV theory way too flavor symmetric and imposes the familiar, nontrivial challenges of UV flavor model building. Thus, writing down a fully realistic model requires a predictive theory of the entire quark Yukawa sector, which is a lofty and important goal; however, for now, we will attempt to factorize issues and return to the task of flowing precisely to the SM in future work. Of course, we must ensure that we can perform this breaking and generate the nontrivial Yukawa matrices y_u, y_d without upsetting our achievement in providing the boundary condition $\bar{\theta}(\Lambda_9) = 0$. In particular, this process includes generating the CP-violating phase in the CKM matrix, as invariantly parametrized by Jarlskog, $\tilde{J} = \text{Im det}([y_u^\dagger y_u, y_d^\dagger y_d])$.

Many approaches to strong CP protect $\bar{\theta}$ from δ_{CKM} in a way that intrinsically relies on the small sizes of some entries in the CKM matrix. In contrast, we describe one possible, general scheme to implement flavor breaking in which $\bar{\theta}$ continues to vanish without relying on the specific structure of the low-energy SM. We use the gauged flavor symmetry to our advantage in recognizing that it provides a good way to generate nontrivial complex structure in the Yukawas while keeping them Hermitian, $y_u^\dagger = y_u, y_d^\dagger = y_d$ (in the canonical UV basis). With the UV PQ symmetry producing $\bar{\theta}(\Lambda_9) = 0$, this method of communicating nontrivial flavor and weak CP violation to the quarks at the scale Λ_3 guarantees that the strong CP problem continues to be solved. We will make a more precise statement in Sec. III C. While a fully realistic understanding of flavor in these theories will require additional work, the structure we describe shows a general scheme for solving the strong CP problem in this framework.

A. The $SU(9)$ unified theory and $\bar{\theta}(\Lambda_9)=0$

An example of a unified theory that provides the $\mathbb{Z}_3$ magnetic monopoles (or $\mathbb{Z}_3$ small instantons) that break the 1-form symmetry of the $SU(3)^2/\mathbb{Z}_3$ gauge theory, hence breaking the noninvertible symmetry, is the embedding in $SU(9)$ color-flavor unification. This embedding is minimal in that it requires no new fermions, being simply a gauging of the global symmetries of the standard model quark fields [38], as evinced in the fermion content given in Table VI. Of course, one could consider alternative ultraviolet theories that also introduce the correct magnetic monopoles, but generically, these theories require additional structure to lift extra fermion species.

In the UV Lagrangian, we explicitly write down the "top" Yukawa, and in a general basis, the UV Lagrangian contains

$$\mathcal{L}_0 = y_t \tilde{H}\mathbf{Q}\bar{\mathbf{u}} + \text{H.c.} + \frac{i\theta_9}{32\pi^2} F\tilde{F}, \tag{3.1}$$

where our notation is $\mathbf{Q}\bar{\mathbf{u}} = \mathbf{Q}^A \bar{\mathbf{u}}_A$ [$A = 1, \ldots, 9$ is an $SU(9)$ index] and $F_{\mu\nu}$ is the field strength of the $SU(9)$ gauge field with its dual defined as usual, $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$. The full UV theory includes a couple more terms, for reasons that we will explain in detail below. For now, we list those terms and give a brief motivation for each of them. In addition to $\mathcal{L}_0$, our UV theory includes the following terms.

(1) $\mathcal{L}_\Phi = |D_\mu \Phi|^2 - V(\Phi)$

Here, Φ is a scalar field transforming in the three-index symmetric representation of $SU(9)$. It is responsible for the breaking $SU(9) \to SU(3)^2/\mathbb{Z}_3$ at a high scale $\langle\Phi\rangle = \Lambda_9$. Even after this breaking, the gauged quark flavor symmetry $SU(3)_H$ implies that the quark Yukawas continue to be flavor symmetric.

(2) $\mathcal{L}_\Sigma = |D_\mu \Sigma_1|^2 + |D_\mu \Sigma_2|^2 - V(\Sigma_1, \Sigma_2)$, $V_{\mathbb{Z}_4}(\Sigma) = \eta_1 \mathrm{Tr}(\Sigma^4) + \eta_2 \mathrm{Tr}(\Sigma^2)^2 + \text{H.c.}$

We introduce two $SU(9)$ adjoint scalars Σ_1 and Σ_2 to further break $SU(3)^2/\mathbb{Z}_3 \to SU(3)_C$, and we assume such a breaking occurs at a scale $\langle\Sigma\rangle = \Lambda_3 \lesssim \Lambda_9$. Since reproducing the observed SM requires not only real entries of 3×3 mass matrices (quark masses and flavor mixings) but also a complex CP-violating phase (δ_{CKM}), we need to introduce CP-violating parameter(s). In our theory, the vevs of $\Sigma_{1,2}$ generate the desired texture, and complex parameters in their potential $V(\Sigma)$ provide necessary CP-violating phases. When the theory respects $\mathbb{Z}_4$ symmetry of Σ (which, however, is not essential for our mechanism to work, as we elaborate on below), the potential takes a simple form as shown above. There, we combined two $\Sigma_{1,2}$ to form a single "complex" adjoint field $\Sigma = \Sigma_1 + i\Sigma_2$, and $\eta_{1,2}$ are two complex parameters. However, note that there are additional terms but with *real* parameters, e.g., $\mathrm{Tr}(\Sigma^\dagger\Sigma)^2$, which we have not written down because they do not play a key role.

(3) $\mathcal{L}_{\rho\chi} = |D_\mu\rho|^2 + i\chi^\dagger\partial\chi + \lambda_d \bar{\mathbf{d}}\rho\chi + a_1\rho\Sigma\rho^\dagger + a_2\rho\Sigma\Sigma\rho^\dagger + \text{H.c.} + \rho(c_1\Sigma^\dagger\Sigma + c_2\Sigma\Sigma^\dagger)\rho^\dagger$

Since the Jarlskog invariant $\tilde{J} \propto \mathrm{Im}\ \det[y_u^\dagger y_u, y_d^\dagger y_d]$ measures the "misalignment" of the up vs down Yukawas, generating the desired flavor structure (both real texture and the CP-violating phase) requires that the up- and down-type quarks are not treated identically. We implement this up-down asymmetry by introducing one completely sterile Weyl fermion χ and one $SU(9)$ fundamental scalar ρ with hypercharge such that only "down-philic" interactions among these and the SM quarks are allowed. [112] Note that ρ is charged under $U(1)_{PQ}$ but does not obtain a vev.

$SU(9)$ representations of these additional fields are summarized in Table VII. In addition, in Table VIII, we list how

TABLE VI. Standard model matter content of the ultraviolet color-flavor unified gauge theory with classical global symmetries.

	$SU(9)$	$U(1)_{\bar{B}}$	$U(1)_{\bar{d}}$
$\mathbf{Q}$	9	+1	0
$\bar{\mathbf{u}}$	$\bar{9}$	−1	0
$\bar{\mathbf{d}}$	$\bar{9}$	0	+1
H	1	0	0

TABLE VII. Additional matter content used to break down to the SM in the infrared. The $SU(9)$-charged fields are all scalars, and χ is a singlet fermion.

	$SU(9)$	$U(1)_{\bar{B}+\bar{d}}$
Φ	165	0
$\Sigma_{1,2}$	80	0
ρ	9	−1
χ	1	0

TABLE VIII. Branching of some $SU(9)$ representations to $(SU(3)_C \times SU(3)_H)/\mathbb{Z}_3$ and $(SU(3)_C \times U(1)_H)/\mathbb{Z}_3$, where non-Abelian representations are in parentheses and Abelian charges are subscripts, with the multiplicity of representations as a prefactor.

$SU(9)$	$(SU(3)_C \times SU(3)_H)/\mathbb{Z}_3$	$(SU(3)_C \times U(1)_H)/\mathbb{Z}_3$
9	(3, 3)	$2(3)_{+1} + (3)_{-2}$
80	$(8,8) + (8,1) + (1,8)$	$5(8)_0 + 2(8)_{+3} + 2(8)_{-3} + 4(1)_0 + 2(1)_{+3} + 2(1)_{-3}$
165	$(10,10) + (8,8) + (1,1)$	$(10)_{-6} + 2(10)_{-3} + 3(10)_0 + 4(10)_{+3} + 4(8)_0 + 2(8)_{-3} + 2(8)_{+3} + (1)_0$

various $SU(9)$ representations decompose upon symmetry breaking.

In the remaining part of this subsection, we show explicitly that strong CP violation is absent in the $SU(9)$ phase of the theory. Subsequent threshold corrections, generation of the CKM phase, and potential renormalization of $\bar{\theta}$ will be discussed in the following sections.

We first note that the top Yukawa explicitly breaks the separate classical global symmetries for $\mathbf{Q}, \bar{\mathbf{u}}$ down to the diagonal $U(1)_{\tilde{B}}$. This process leaves two classical $U(1)$ s, which, at the quantum level, are subject to ABJ anomalies $SU(9)^2 U(1)_{\tilde{B}} = SU(9)^2 U(1)_{\bar{d}} = 1$; thus, they should be arranged into the familiar $SU(9)$-anomaly-free baryon number $B = \tilde{B} - \bar{d}$. The other direction is a flavor-unified Peccei-Quinn symmetry, which we take as $\tilde{B} + \bar{d}$. [113] Since the theory is asymptotically free, in the far UV, the violation of the Peccei-Quinn symmetry by the anomaly becomes arbitrarily weak, $\exp(-2\pi/\alpha) \to 0$. We assume this is a good symmetry of the UV and its only breaking is by these instanton effects. [114] Then, at the classical level, the down Yukawa y_d is forbidden by the Peccei-Quinn symmetry.

In a general basis, $y_t = |y_t| e^{i\theta_t}$ is some complex number, and the gauge theory has a phase $\theta = \theta_9$. We may perform a field redefinition $\bar{\mathbf{u}} = \bar{\mathbf{u}}' e^{-i\theta_t}$ to make the up Yukawa real, bearing out the general EFT understanding of using spurions to count physical phases (see, e.g., Ref. [115]). The $\bar{\mathbf{u}}$ rotation is anomalous, so this rotation also changes the topological density term to $\theta = \theta_9 - \theta_t$. Since the down quark is classically massless, we can then perform a rotation $\bar{\mathbf{d}} = \bar{\mathbf{d}}' e^{-i(\theta_9 - \theta_t)}$ to manifestly remove the dependence of the Lagrangian on the topological density term. Thus, there is a "canonical" basis in which the theta angle is absent and the masses are all real; i.e., there is no strong CP violation in the $SU(9)$ phase with a good $U(1)_{PQ}$.

Quantum mechanically, however, the Peccei-Quinn symmetry is broken by the instantons of $SU(9)$, and we will discuss this effect in Sec. III B along with the breaking of $SU(9)$, which dictates the small instanton scale. We give additional details in Appendix C.

B. $SU(9)$ breaking and instanton effects

The first step of symmetry breaking to $SU(3)^2/\mathbb{Z}_3$ may be achieved by the condensation of a three-index symmetric Φ^{ABC}, which obtains a vev

$$\langle \Phi^{ABC} \rangle = \Lambda_9 \varepsilon^{abc} \varepsilon^{ijk}, \tag{3.2}$$

where the $SU(9)$ indices are reinterpreted as multi-indices under the two $SU(3)$ factors, which are manifestly preserved, $A \in \{1, 2, \ldots, 8, 9\} \leftrightarrow ai \in \{11, 12, \ldots, 23, 33\}$. The fundamental 9 branches to the bifundamental $3 \otimes 3$ as

$$\bar{u}^A = \begin{pmatrix} \bar{u}^1 \\ \cdots \\ \bar{u}^9 \end{pmatrix} = \begin{pmatrix} \bar{u}^{ru} & \bar{u}^{gu} & \bar{u}^{bu} \\ \bar{u}^{rc} & \bar{u}^{gc} & \bar{u}^{bc} \\ \bar{u}^{rt} & \bar{u}^{gt} & \bar{u}^{bt} \end{pmatrix} = \bar{u}^{ai}, \tag{3.3}$$

and one may usefully envision this as an "outer product" decomposition into a matrix of the quark colors and flavors. The different $SU(3)$ factors now act as left or right matrix multiplication, and the nontrivial global structure is seen simply because a left multiplication by an element of the center, $\omega \mathbb{1}_C$ with ω a cube root of unity, commutes through and can cancel against a right multiplication by $\omega^{-1} \mathbb{1}_H$. In other words, for some general center rotations,

$$\begin{pmatrix} e^{i\frac{2\pi}{3}n_h} & 0 & 0 \\ 0 & e^{i\frac{2\pi}{3}n_h} & 0 \\ 0 & 0 & e^{i\frac{2\pi}{3}n_h} \end{pmatrix} \begin{pmatrix} \bar{u}^{ru} & \bar{u}^{gu} & \bar{u}^{bu} \\ \bar{u}^{rc} & \bar{u}^{gc} & \bar{u}^{bc} \\ \bar{u}^{rt} & \bar{u}^{gt} & \bar{u}^{bt} \end{pmatrix} \begin{pmatrix} e^{i\frac{2\pi}{3}n_c} & 0 & 0 \\ 0 & e^{i\frac{2\pi}{3}n_c} & 0 \\ 0 & 0 & e^{i\frac{2\pi}{3}n_c} \end{pmatrix} = e^{i\frac{2\pi}{3}(n_h + n_c)} \bar{u}^{ai}, \tag{3.4}$$

and all the matter in the theory is invariant along the $n_h = -n_c$ direction since all the other irreducible representations are contained in products of the fundamental and antifundamental. The Φ branches as $165 \to (1,1) + (8,8) + (10,10)$, where the 8's are the adjoints, the 10's are the three-index symmetric tensor, and the entire (8,8) is eaten by the gauge bosons. The SM matter fields are as in Table IV.

Since the $SU(9)$ theory is asymptotically free, the dominant contribution to the 't Hooft vertices arises at the scale Λ_9 where the breaking $SU(9) \to SU(3)^2/\mathbb{Z}_3$ occurs. Note also that, across Λ_9, the gauge couplings are nontrivially matched as $1/g_3^2 = 3/g_9^2$, as there is a nontrivial "index of embedding" [22,98,116] of $SU(3)^2/\mathbb{Z}_3$ into $SU(9)$. Nontrivial index of embedding means the following. Given a breaking of a gauge group $G \to H$, if the index of embedding is greater than 1, then not all of the G-instanton effects are captured by unbroken H instantons. Here, the

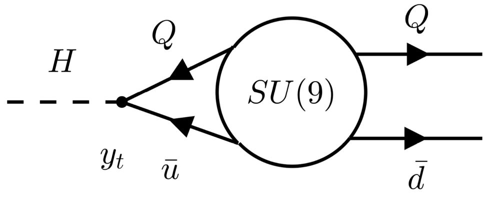

FIG. 2. The 't Hooft vertex in the $SU(9)$ theory by which instantons generate the down-type Yukawa from the up-type Yukawa.

index of embedding is 3. Thus, there are $SU(9)$ instantons that appear to be $\mathbb{Z}_3$ fractional instantons of $SU(3)^2/\mathbb{Z}_3$ theory, as our analysis in Sec. II showed could occur. These instantons are precisely what we need to explicitly break the noninvertible symmetries. The nontrivial index of embedding not only increases the size of UV instanton effects but also decreases the number of legs of the 't Hooft vertex. Together, these effects are crucial for our solution.

The instantons of the $SU(9)$ theory, in violating the anomalous global symmetries, generate the 't Hooft vertex of Fig. 2. Since we avoided adding any additional $SU(9)$-charged fermions, the 't Hooft vertices include only the SM fermion zero modes. However, the additional charged scalars will affect the instanton density, and we defer a fuller discussion of the sizes of the effects of 't Hooft vertices to Appendix C. Here, we content ourselves with the rough result that instantons generate a down-type Yukawa that violates $U(1)_{\tilde{B}+\bar{d}}$ PQ symmetry by two units. Flowing down in energies and momentarily staying at some scale $\Lambda > \Lambda_9$, instantons begin to generate

$$\mathcal{L}(\Lambda) \sim y_t H\mathbf{Q}\bar{\mathbf{u}} + y_t^\star e^{i\theta_9} e^{-\frac{2\pi}{\alpha_9(\Lambda)}} H\mathbf{Q}\bar{\mathbf{d}} + \text{H.c.} + \frac{i\theta_9}{32\pi^2} F\tilde{F}. \tag{3.5}$$

We see that a color-flavor symmetric down-type Yukawa is generated with a rough size,

$$y_d \sim y_t^\star e^{i\theta_9} e^{-\frac{2\pi}{\alpha_9(\Lambda)}}. \tag{3.6}$$

A key point is that such an instanton-induced Yukawa comes with just the right phase to ensure

$$\bar{\theta} = \arg e^{-i\theta_9} \det y_u y_d = -\theta_9 + \arg |y_t|^2 e^{i\theta_9} = 0, \tag{3.7}$$

where, because of the color-flavor gauge symmetry, the Yukawas are simply numbers. This seeming conspiracy among phases is guaranteed by the good PQ symmetry of the UV. In the canonical basis, the Yukawas remain real.

Now, moving to the theory at the matching scale, in a general basis, we have

$$\mathcal{L}(\Lambda_9) \sim y_t H\mathbf{Q}\bar{\mathbf{u}} + y_t^\star e^{i\theta_9} e^{-\frac{2\pi}{3\alpha_s(\Lambda_9)}} H\mathbf{Q}\bar{\mathbf{d}} + \text{H.c.} + \frac{i3\theta_9}{32\pi^2}(G\tilde{G} + K\tilde{K}), \tag{3.8}$$

where $\alpha_s(\Lambda_9) = \alpha_9(\Lambda_9)/3$ is the QCD coupling evolved from the infrared up to the Λ_9 scale. Here, G and K are the $SU(3)_C$ and $SU(3)_H$ gauge field strengths, respectively. The nontrivial matching of the theta angles accounts for the Yukawas being upgraded from single numbers to 3×3 matrices [from the perspective of $SU(3)_C$] and ensures

$$\bar{\theta} = -3\theta_9 + \arg\det |y_t|^2 e^{i\theta_9} = 0, \tag{3.9}$$

which, again in the canonical basis, reduces to the statement that the Yukawas are all real and thus have real eigenvalues. This finding is the core of the massless quark solution to the strong CP problem [85–88].

In contrast to our generation of Dirac neutrino masses from the charged lepton masses [3], here the required suppression from the top to the bottom Yukawa is not so large, $y_b/y_t \sim 1/40$. From the naive one-instanton Eq. (3.8), we can estimate $\alpha_s(\Lambda_9) \simeq 0.57$, and we comment on the effects of quadratic fluctuations around this solution in Appendix C. In the end, it is difficult to obtain a reliable analytic estimate of the instanton effects for our $SU(9)$ theory. Just as with the original massless up-quark solution, lattice simulations will be needed in order to determine precisely how well this works. However, we are aided by the gauge coupling increasing due to the nonminimal index of embedding at the $SU(9)$ scale; we also have a natural model-building handle to slow down or reverse its one-loop running through colored particles with masses below Λ_9, which will be discussed further in Sec. III D.

C. Flavor breaking and keeping $\bar{\theta} \approx 0$

Having described a color-flavor unified theory that guarantees $\bar{\theta} = 0$ in the ultraviolet, we need to understand how $\delta_{\text{CKM}} \sim \mathcal{O}(1)$ may appear without spoiling it. As discussed above, making use of gauged flavor symmetry forces us to confront the generation of the non-flavor-symmetric SM Yukawa sector. Here, we pursue the simplest possibility of a single, further symmetry-breaking step $(SU(3)_C \times SU(3)_H)/\mathbb{Z}_3 \to SU(3)_C$, in which we break the horizontal symmetry all at once at a lower scale, $\Lambda_3 < \Lambda_9$. Our strategy to not destabilize our UV achievement will be to communicate flavor and CP breaking to the SM quarks in a way that keeps the Yukawas Hermitian. Hermiticity of the quark Yukawas has been used in different ways for the strong CP problem before—for example, in parity-symmetric theories [117] or with supersymmetry [62,63] or in an effective 2HDM [118]—but our usage will be quite novel. In a theory of gauged flavor, it will be a natural possibility as we will see below.

This separation of the scales at which the instantons generate noninvertible symmetry violation and at which the flavor structure is generated is a possibility in this theory that differs from the theories considered in Refs. [93,94]. [119] Generally, it could aid in keeping $\bar{\theta}$ small, but here we will describe a mechanism to generate flavor that automatically preserves $\bar{\theta} = 0$ without utilizing this structural possibility. As seen in the prior sections, we begin with Yukawas proportional to the identity matrix $y_u, y_d \propto \mathbb{1}$. Generating the nontrivial flavor structure, especially flavor hierarchies, using higher-dimensional operators suppressed by powers of Λ_3/Λ_9 means that there is a tension between reproducing the SM flavor structure and having a large ratio of scales in this scheme [120]. That is, we will not ask for any large hierarchy between these two scales, nor will we discuss a predictive theory of their origins.

There may be many choices of how to break $SU(3)_H$ and match onto the SM. As our symmetry-breaking sector, we choose two $SU(3)_H$ adjoint scalar fields $\Sigma_{1,2}$, which, together, can entirely break $SU(3)_H \to \varnothing$. We find it useful to join these scalar fields together into the "complex adjoint" $\Sigma \equiv \Sigma_1 + i\Sigma_2$, which can be seen merely as an accounting measure to keep track of their would-be $SO(2) \simeq U(1)_\Sigma$ global symmetry. Their nonzero commutator $[\Sigma_1, \Sigma_2] = [\Sigma^\dagger, \Sigma]/(2i)$ is required to fully break the $SU(3)_H$ symmetry. In addition, as the SM Jarlskog invariant is written in terms of a commutator of flavor spurions, it will be proportional to this single, nonvanishing commutator in this model.

With only the SM fermions, there are no renormalizable interactions allowed with Σ, and its breaking of $SU(3)_H$ is communicated to the quarks solely through the broken $SU(9)$ gauge bosons. [121] As these flavor-breaking effects must include the generation of the CP-violating δ_{CKM}, we must ensure that this scalar sector can break CP. Indeed, the most general potential for Σ includes many CP-violating phases, and for simplicity, we can find a more tractable potential by imposing a $\mathbb{Z}_4$ symmetry,

$$V_{\mathbb{Z}_4}(\Sigma) = \eta_1 \mathrm{Tr}(\Sigma^4) + \eta_2 \mathrm{Tr}(\Sigma^2)^2 + \mathrm{H.c.} + \text{terms with real coefficients}, \tag{3.10}$$

where we have left off the terms that do not break $U(1)_\Sigma$. This potential has a single CP-odd phase, which is captured by the field-redefinition-invariant $\eta_1^\dagger \eta_2$. We assume it has some random complex phase, explicitly breaking CP. In this simplified scenario, there is no spontaneous violation of CP when Σ obtains a vev because η_1 and η_2 both have charge -4 under the spurious $U(1)_\Sigma$ symmetry [122]. In general, without imposing $\mathbb{Z}_4$, the Σ potential will both explicitly and spontaneously violate CP. Either way is fine; the mechanism we now describe works no matter how CP

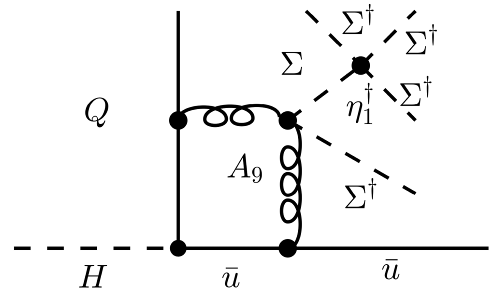

FIG. 3. The Λ_9 threshold contribution to the up-type Yukawas, which eventually gives flavor and CP breaking proportional to $\alpha_9 \eta_1^\dagger \langle \Sigma^\dagger \rangle_b^{4a}/(4\pi)$. The Hermitian conjugate contribution is generated by the same diagram with an η_1 insertion on the conjugate scalar leg; thus, together, they yield a Hermitian up-type Yukawa. Here, Σ is the $SU(3)_H$ adjoint.

violation appears in Σ's potential, and we will explore the spontaneous case further in forthcoming work [123].

Despite the fact that we now have flavor and CP breaking, in this model, $\bar{\theta} = 0$ continues to hold because the effects of Σ (both in producing the Yukawa texture and in CP violation) are transferred to the SM fields only through the $SU(9)$ heavy-gauge bosons. As a result, for any coupling that one could attach to an external Σ leg to obtain a complex contribution to a Yukawa coupling, there is also always a Hermitian conjugate coupling to be attached to the $\Sigma^\dagger$ leg. See, for example, Fig. 3. As a whole, our mechanism works in a way such that corrections to the Yukawas always leave them Hermitian. Note that this argument applies just as well to the RG evolution during the $(SU(3)_C \times SU(3)_H)/\mathbb{Z}_3$ phase when the Σ fields appear only in closed loops—while there are many-loop diagrams proportional to $\eta_1^\dagger \eta_2$, there are compensating diagrams proportional to $\eta_2^\dagger \eta_1$ that sum up to real Yukawas in the $(SU(3)_C \times SU(3)_H)/\mathbb{Z}_3$ phase when they are still just numbers.

Then, below the $SU(3)$-breaking scale Λ_3, the basic structure of the up-type Yukawas, for example, is given by

$$\begin{aligned}(y_u)^a{}_b \sim y_t \Bigg(& \mathbb{1}^a{}_b + \frac{\alpha_9}{(4\pi)} \frac{\{\Sigma^\dagger, \Sigma\}^a_b}{2\Lambda_9^2} \\ & + \text{other terms with real coefficients} \\ & + \frac{\alpha_9}{(4\pi)} \frac{\eta_1^\dagger (\Sigma^{\dagger 4})^a_b + \eta_2^\dagger \mathrm{Tr}(\Sigma^{\dagger 2})(\Sigma^{\dagger 2})^a_b}{\Lambda_9^4} \\ & + \frac{\alpha_9}{(4\pi)} \frac{\eta_1 (\Sigma^4)^a_b + \eta_2 \mathrm{Tr}(\Sigma^2)(\Sigma^2)^a_b}{\Lambda_9^4} + \ldots \Bigg),\end{aligned} \tag{3.11}$$

where we have only given some naive power counting for a notion of the size of these effects, and it should be understood that these are vevs of Σ. The anticommutator $\{\Sigma^\dagger, \Sigma\}/2 = \Sigma_1^2 + \Sigma_2^2$ is the structure hat appears from the gauge interactions, and these must respect the $U(1)_\Sigma$

spurious "flavor" symmetry. This anticommutator is the leading correction to the flavor structure which is manifestly real, all of which we group on the top line. In the second line, we give the first corrections through which a complex phase enters the Yukawas in the simplified scenario where the $\mathbb{Z}_4$ symmetry of Σ controls the potential.

So far, the corrections consist of a sum over diagrams attaching a vertex with a complex coefficient to the Σ line and from attaching its conjugate vertex to the $\Sigma^\dagger$ line. Then, while the generated Yukawas are no longer real, they remain Hermitian, which suffices to guarantee $\det y_u, \det y_d \in \mathbb{R}$, and so $\bar{\theta} = 0$. Note also that, while in the unbroken phase the CP violation is necessarily invariantly parametrized only by $\eta_1^\dagger \eta_2$, at low energies after Σ obtains a vev, we integrate out its fluctuations, and it appears only as an external source. Thus, the combination $\eta_1^\dagger \mathrm{Tr}\langle\Sigma\rangle^{\dagger 4} + \eta_2^\dagger \mathrm{Tr}(\langle\Sigma\rangle^{\dagger 2})^2$ can also invariantly parametrize CP violation.

However, while we now have complex, $SU(3)_H$-violating Yukawas, this setup cannot yet generate the CKM CP angle. The issue is that the $SU(9)$ dynamics affect the up and down quarks symmetrically, whereas a nonvanishing CKM phase appears from a mismatch in the form of the Yukawa couplings, $\sin\delta_{\mathrm{CKM}} \propto \mathrm{Im}\ \det([y_u^\dagger y_u, y_d^\dagger y_d])$. Thus, we must introduce another ingredient to skew the Yukawas apart while not upsetting the solution.

In particular, we can introduce some fields that interact only with, say, the down quark and not the up quark. This is simplest if it does not allow for any new CP phases in operators containing quarks nor introduce any new color-flavored fermions which would appear in our 't Hooft vertices. With two new fields ρ and χ, a new Yukawa is allowed, where one allocation of quantum numbers is for the scalar ρ to be a down squark and the fermion χ to be sterile. The scalar can furthermore couple directly to Σ,

$$\begin{aligned}\mathcal{L}_{\rho\chi} \supset\ & \lambda_d \bar{\mathbf{d}}\rho\chi + a_1\rho\Sigma\rho^\dagger + a_2\rho\Sigma\Sigma\rho^\dagger + \mathrm{H.c.} \\ & + \rho(c_1\Sigma^\dagger\Sigma + c_2\Sigma\Sigma^\dagger)\rho^\dagger,\end{aligned} \tag{3.12}$$

where we have suppressed indices to avoid notational clutter, e.g., $\rho\Sigma\Sigma\rho^\dagger = \rho_a\Sigma^a_b\Sigma^b_c(\rho^\dagger)^c$. These "down-philic" interactions generate a loop correction to the down-type Yukawas that is not present for the up-type Yukawas, which allows a CKM phase to be generated, as we now discuss. We can use a χ rotation to make λ_d real, and $c_{1,2}$ are real by self-Hermiticity of the operators. Note that $a_{1,2}$ are set to zero if the $\mathbb{Z}_4$ is imposed or, in general, have complex phases and lead to further field-redefinition-invariant CP-odd parameters such as $a_1^2 a_2^\dagger$ or $\eta_1^\dagger a_2^2$. In either case, the interactions of ρ to Σ allow further flavor violation to be communicated to the quarks in a way that does not upset our mechanism: Σ couples to $\rho^\dagger\rho$, such that it will always enter in a Hermitian manner. See Fig. 4.

Overall, this mechanism works by having all CP-violating phases in the scalar sector and communicating them to the SM quark sector via a bosonic mediation. In particular, the interactions of the Σ fields with the mediators (who will transfer the flavor and CP breaking to the SM quarks) are always Hermitian in the mediator fields, though not necessarily in the Σ fields. Then, the requirement that the Lagrangian is Hermitian itself ensures that, for any diagram with possibly complex phases, there always exists a conjugate diagram where the $U(1)_\Sigma$ charges are all reversed, and one obtains the complex conjugate phase. This case would not be true if the CP-violating phases were directly coupled to the SM fermions since conjugating the diagram would then require charge conjugating the fermion legs, resulting in a different diagram from the original one. Instead, here we automatically obtain a sum over all Σ and $\Sigma^\dagger$ source insertions when computing the Yukawa corrections, which keeps them Hermitian.

Now, let us examine the form of the Yukawas and the CP-violating phases we have generated. Recall that, aside from the strong CP angle itself, the only field-redefinition-invariant CP-odd parameter in the theory of quarks is the Jarlskog invariant,

$$\tilde{J} = \mathrm{Im}\ \det([y_u^\dagger y_u, y_d^\dagger y_d]). \tag{3.13}$$

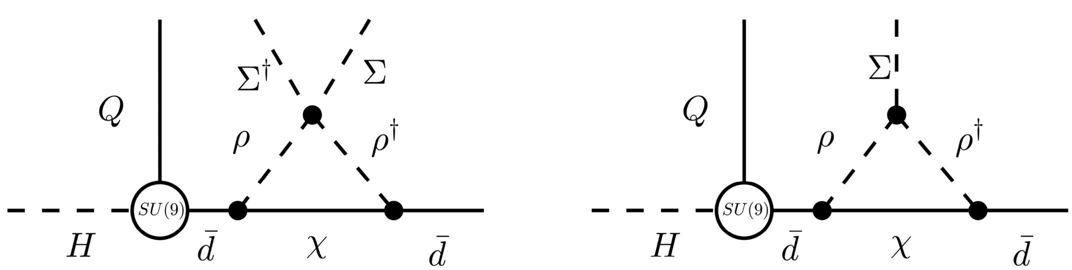

FIG. 4. Threshold corrections to the down-type Yukawas. The left diagram preserves $\mathbb{Z}_4$ and gives a flavor-breaking but CP-preserving contribution proportional to $|\lambda_d|^2 c_1 \langle\Sigma^\dagger\Sigma\rangle^a_b$. The right diagram is present in the general $\mathbb{Z}_4$-violating case, and it gives flavor and CP breaking proportional to $|\lambda_d|^2 a_1 \langle\Sigma\rangle^a_b$. Again, the Hermitian conjugate contribution is generated by the same diagram, with an $a_1^\dagger$ insertion of $\Sigma^\dagger$, so they sum to yield a Hermitian down-type Yukawa. These ρ-mediated corrections appear solely for the down-type quarks.

This parameter is small, $\tilde{J} \simeq [(m_c)/(m_t)][(m_s)/(m_b)]J \sim 5 \times 10^{-9}$.

Since the only object breaking the flavor symmetry is Σ, the CKM phase must depend on $[\Sigma, \Sigma^\dagger]$ as the only nonvanishing commutator. Now, working even more schematically just to see the structure, in the $\mathbb{Z}_4$-conserving case, we have (see the left panel of Fig. 4)

$$y_u \sim \mathbb{1} + \{\Sigma^\dagger, \Sigma\} + (\eta\Sigma^4 + \eta^\dagger\Sigma^{\dagger 4}) + \ldots, \tag{3.14}$$

$$y_d \sim \mathbb{1} + c\Sigma^\dagger\Sigma + \{\Sigma^\dagger, \Sigma\} + (\eta\Sigma^4 + \eta^\dagger\Sigma^{\dagger 4})\ldots, \tag{3.15}$$

where we have only written down enough terms to see both that CP violation enters (through the η complex self-couplings of Σ) and that the two structures differ (due to the c couplings of ρ to Σ). Moreover, we have absorbed Λ_9^{-1} into the scalar field, and we have left off most of the constants, including the overall proportionality factors $y_u \propto y_t$ and $y_d \propto y_b$, where $y_b \sim y_t^\star e^{i\theta_9} e^{-2\pi/\alpha_9(\Lambda_9)}$. As discussed above, we can choose a basis where both are real numbers, and for convenience, we do so. To analyze the Jarlskog invariant, it can be useful to split $y_d = r y_u + r\Delta y$ into a piece that is simply a real rescaling ($r \in \mathbb{R}$) of the y_u structure (so will manifestly commute) and an extra piece, which, for us, is given by the extra effects of the ρ and χ interactions. Here, $r \sim e^{-2\pi/\alpha_9(\Lambda_9)}$ and $\Delta y \sim c\Sigma^\dagger\Sigma$, with $c \in \mathbb{R}$. Then, we have

$$\tilde{J} = \mathrm{Im}\ \det\left(r^2[y_u^2, \{y_u, \Delta y\} + \Delta y^2]\right) \tag{3.16}$$

$$\supsetneq \mathrm{Im}\ \det\left(4r^2[\eta\Sigma^4 + \eta^\dagger\Sigma^{\dagger 4}, c\Sigma^\dagger\Sigma]\right) + \ldots, \tag{3.17}$$

where, in the second line, we only show the lowest-order structure that can contribute to weak CP violation assuming $\mathbb{Z}_4$ invariance. In general, without $\mathbb{Z}_4$, this could come in at a lower order. The r^2 dependence reflects the fact that $\tilde{J}$ is proportional to y_d^2, while c dependence shows that $\tilde{J} = 0$ without the up-down asymmetry factor. We can rewrite this case in terms of the commutator $[\Sigma, \Sigma^\dagger]$ as

$$\begin{aligned}\tilde{J} \propto \mathrm{Im}\ \det(&\eta([\Sigma,\Sigma^\dagger]\Sigma^4 + \Sigma[\Sigma,\Sigma^\dagger]\Sigma^3 + \Sigma^2[\Sigma,\Sigma^\dagger]\Sigma^2 + \Sigma^3[\Sigma,\Sigma^\dagger]\Sigma)\\ &- \eta^\dagger(\Sigma^\dagger[\Sigma,\Sigma^\dagger]\Sigma^{\dagger 3} + \Sigma^{\dagger 2}[\Sigma,\Sigma^\dagger]\Sigma^{\dagger 2} + \Sigma^{\dagger 3}[\Sigma,\Sigma^\dagger]\Sigma^\dagger + \Sigma^{\dagger 4}[\Sigma,\Sigma^\dagger])).\\ = \mathrm{Im}\ \det(&\eta([\Sigma,\Sigma^\dagger]\Sigma^4 + \Sigma[\Sigma,\Sigma^\dagger]\Sigma^3 + \Sigma^2[\Sigma,\Sigma^\dagger]\Sigma^2 + \Sigma^3[\Sigma,\Sigma^\dagger]\Sigma) - \mathrm{H.c.}).\end{aligned} \tag{3.18}$$

This result will be generically nonvanishing so long as $\eta \notin \mathbb{R}$ and $[\Sigma, \Sigma^\dagger] \neq 0$, which is also necessary for $\langle\Sigma\rangle$ to properly break the horizontal symmetry.

Therefore, we have succeeded in generating the weak CP phase of the quarks while keeping the strong phase vanishing due to the Hermiticity. In a more general case that does not preserve $\mathbb{Z}_4$, the power counting can turn out differently. For example, because we can rely on CP violation in the coupling $\rho^\dagger(a\Sigma + a^\dagger\Sigma^\dagger)\rho$ rather than in only Σ self-couplings. This will yield

$$\begin{aligned}\tilde{J} \propto \mathrm{Im}\ \det(&a^\dagger([\Sigma,\Sigma^\dagger]\Sigma^\dagger + \Sigma^\dagger[\Sigma,\Sigma^\dagger])\\ &- a(\Sigma[\Sigma,\Sigma^\dagger] + [\Sigma,\Sigma^\dagger]\Sigma)).\end{aligned} \tag{3.19}$$

While still suppressed by the gauge coupling, the bottom Yukawa, and two loop factors as before, we have less suppression now, by a factor of about $(\Lambda_3/\Lambda_9)^3$.

Let us review the model to discuss what values of $\bar{\theta}$ these models predict. In the $SU(9)$ phase above Λ_9, there is no strong CP violation as a result of $U(1)_{PQ}$. At Λ_9, instantons generate PQ violation but ensure $\bar{\theta} = 0$. In the $(SU(3)_C \times SU(3)_H)/\mathbb{Z}_3$ phase, the Yukawas are simply numbers, so there is no other CP-odd quark parameter. The RG evolution in this phase and the matching at Λ_3 when Σ obtains a vev both have a structure that generates complex, but Hermitian Yukawas, keeping $\bar{\theta} = 0$. Only below Λ_3, after integrating out Σ, is the CKM phase δ_{CKM} present to renormalize $\bar{\theta}$. However, now we are back to the SM field content, so this renormalization is minuscule and protected by the SM structure, with the finite renormalization producing only $\bar{\theta} \sim 10^{-16}$, as estimated by Ellis and Gaillard [45].

Before concluding this section on $SU(3)_H$ breaking, let us emphasize what we have achieved. Often, ensuring that models designed to solve the strong CP problem do not generate unacceptably large $\bar{\theta}$ relies explicitly on the small entries of the SM Yukawa matrices. Recall that this is the case in the SM itself, as many loops are needed to generate $\bar{\theta}$ from δ_{CKM}, so it is sensible to model build toward that same conclusion. One such example is the flavor-deconstructed massless quark solution of Ref. [93]. However, in this work, we start with the flavor symmetry fully gauged, and we need to generate the nontrivial Yukawas. Then, utilizing small quark mixing angles to keep $\bar{\theta}$ small is more challenging for us since we would first need to construct a predictive theory of flavor, i.e., specify the full texture of $\langle\Sigma\rangle$ and scalar potential compatible with observations.

Instead, we propose a mechanism that factorizes the issue of strong CP from the precise details of the SM flavor structure. Namely, our tactic is to describe a way to generate nontrivial flavor structure and weak CP violation that does not automatically produce nonvanishing $\bar{\theta}$. When the flavor structure of $\langle\Sigma\rangle$, $\langle\Sigma^\dagger\rangle$ is communicated to the SM quarks in the way described above, the Yukawas continue to be Hermitian and $\bar{\theta} = 0$ holds, no matter what the flavor structure is. In other words, this solution would continue to work even if the quark sector were fully anarchic and the Yukawas were described only with $\mathcal{O}(1)$ numbers. Furthermore, we note that this mechanism has an immediate application to a flavorful theory of spontaneous *CP* violation, which we will report on in forthcoming work [123].

D. Scales and running

We want to find the scale Λ_9 at which color unification occurs, which is dictated by the gauge coupling $\alpha_9(\Lambda_9)$ being the right size to generate the bottom Yukawa from the top Yukawa. Since we need relatively large gauge coupling, this scale is subject to theoretical uncertainty and lattice simulations are needed to accurately determine the size of instanton effects. Furthermore, the evolution of the gauge coupling is dictated by the entire charged matter spectrum, which depends both on the representations we have added to implement flavor breaking and also on their masses, so, in principle, there is a lot of freedom.

Across the breaking scale Λ_9, $SU(3)_C$ is embedded nontrivially into $SU(9)$ with an index of embedding of 3, such that the gauge coupling is rescaled as

$$\alpha_9(\Lambda_9) = 3\alpha_s(\Lambda_9). \tag{3.20}$$

This relative factor of 3 strengthens the small $SU(9)$ instanton effects with respect to those at lower energies, which is one reason they may achieve some qualitatively new effects. However, as discussed above, the gauge coupling needed is large, and there may need to be additional running between the $SU(9)$-breaking scale Λ_9 and the scale Λ_3 at which we break to $SU(3)_C$.

Above Λ_9, the $SU(9)$ theory should be asymptotically free, which is easily achieved because of the large number of colors. At the scale Λ_3, we match onto the SM, and heavy-gauge bosons generate flavor-changing four-Fermi operators that are tightly constrained; thus, we must have $\Lambda_3 \gtrsim 1000$ TeV [124,125]. In between these scales, the theory must switch to being IR-free such that the coupling grows into the UV by the beta function

$$\alpha_s^{-1}(\Lambda_9) \simeq \alpha_s^{-1}(\Lambda_3) + \frac{\beta_3}{2\pi}\log\left(\frac{\Lambda_9}{\Lambda_3}\right), \tag{3.21}$$

$$\beta_3 = \left(\frac{11}{3}N_c - \frac{4}{3}n_f I_f - \frac{1}{3}n_s I_s - \frac{1}{6}n_r I_r\right), \tag{3.22}$$

where the charged matter consists of n_f Dirac fermions, n_s complex scalars, and n_r real scalars with Dynkin indices I_f, I_s, and I_r, respectively. The SM has $n_f = 2N_g$ fundamental Dirac fermions. We normalize the $SU(N)$ generators so that the Dynkin index for the fundamental representation is $1/2$, giving $I_f = 1/2$. As outlined in Table VII, our breaking sector has added $2 \times (1 + 8)$ adjoint scalars with $I_r = N_c$ coming from Σ, ten three-index symmetric scalars with $I_s = 15/2$ coming from the components of Φ that were not eaten, and N_g fundamental scalars with $I_s = 1/2$ from ρ. Thus, we have far more than enough colored matter to overpower the gluonic contribution if these scalars, for some reason, do not become massive until a scale below Λ_9. Then, one only needs an extremely mild hierarchy for appreciable running to take place, with $\beta_3 = -55/2$.

From the bottom up, we have precisely measured the low-energy value of the strong coupling, and the PDG gives the world average $\alpha_s(M_Z) \approx 0.118$ [43]. From the scales M_Z up to Λ_3, we have the SM degrees of freedom, which contribute to a beta function $\beta_{QCD} = (11 - 2/3n_f)$, with $n_f = 5$ below $m_t \approx 173$ GeV and $n_f = 6$ above.

For simplicity, we consider a scenario where all of the new colored scalar degrees of freedom obtain masses only at $\Lambda_3 = 10^6$ GeV, and in Fig. 5, we plot the relationship between Λ_9 and $\alpha_9(\Lambda_9) = 3\alpha_s(\Lambda_9)$ by integrating the one-loop RGE. In this scenario, depending on the size of α_9 needed for instanton effects to be large enough, the UV unification scale Λ_9 need be no more than a loop factor above the flavor-breaking scale Λ_3. Suppressing the colored scalar masses relative to Λ_9 is then only a mild tuning. Of course, one could also consider increasing Λ_3, or consider the scalars obtaining an intermediate mass, $\Lambda_3 < M < \Lambda_9$, or keeping fewer scalars light and having larger scale separation; we leave further consideration for future work aimed at more realistic phenomenology.

E. No quality problem

Axion solutions to strong *CP* famously suffer a severe quality problem that their $U(1)_{PQ}$ symmetries are easily

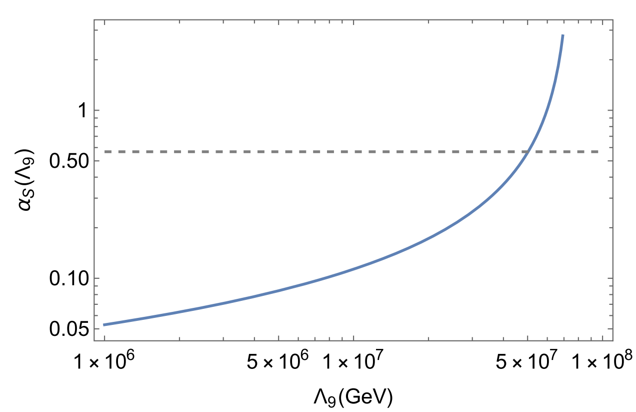

FIG. 5. The strong coupling at the matching scale in a simplified scenario where the colored scalars all have a common mass, $\Lambda_3 = 10^6$ GeV. At Λ_9, the gauge coupling increases by the index of embedding $\alpha_9(\Lambda_9) = 3\alpha_S(\Lambda_9)$, but above Λ_9, the theory is again asymptotically free. For some very rough guidance, the dashed line is the naive estimate of Eq. (3.8) to generate the bottom Yukawa.

destabilized. While it is true that we also require a good $U(1)_{PQ}$ symmetry, our model is not plagued by a quality problem. There are two issues to discuss. First, the purely field theoretic issue is how to clearly understand what it means to impose an anomalous symmetry on a classical action. While the program of generalized symmetries may lead us to better understand the sense in which we can think of instanton effects as spurions for the anomalous symmetry, this issue is not yet entirely clear.

A further and more general concern is that quantum gravitational effects are expected to violate any global symmetries, such that demanding an exact global symmetry in the UV seems theoretically unsound. We still have little understanding of precisely how such violation occurs, but one is at least motivated to consider the effects of Planckian operators made out of the fields in our theory that preserve only gauge symmetries and violate global ones. The potentially disastrous implications for axion models were first discussed in Refs. [126–129].

This challenge is far worse for the axion solution than for the massless quark solution for the following reasons:

(1) The presence of a scalar field ϕ charged under $U(1)_{PQ}$ symmetry allows global-symmetry-violating terms of any dimension, e.g., $\mathcal{L} \supset c_n M_{\rm Pl}^{4-n}\phi^n$, which could conceivably be generated by quantum gravitational effects.

(2) Astrophysical constraints on weakly coupled particles interacting with SM quarks impose a lower limit on the "axion decay constant," $f_a = \langle\phi\rangle$, $f_a \gtrsim 10^8$ GeV [130].

(3) The gravitational, global-symmetry violating operators must be subleading to the QCD-sized potential, which must localize $a \sim -\theta$ to very high accuracy. Thus, the struggle is between $\Lambda_{QCD}^4(1-\cos(a+\theta))$ and $f_a^4(f_a/M_{Pl})^{n-4}(1-\cos(na+\varphi_n))$, where $\varphi_n = \arg c_n$, and the minimum of $a+\theta$ must end up smaller than $\bar{\theta}$.

The symmetry-violating effects of operators generated by quantum gravity should vanish in the limit $M_{\rm pl} \to \infty$ where gravity is turned off. With no other scales around, this justifies the idea that we should consider Planck-suppressed irrelevant operators, as we cannot write coefficients of marginal or relevant terms that have the correct limiting behavior. Considering the least-suppressed irrelevant operator, if the dimension-5 operator has a random $\mathcal{O}(1)$ phase, its coefficient must satisfy

$$|c_5| \lesssim 10^{-35}\left(\frac{\bar{\theta}}{10^{-10}}\right)\left(\frac{10^8\ \text{GeV}}{f_a}\right)^5, \qquad (3.23)$$

very clearly violating any naturalness principle and stretching the plausibility of these models unless there is extra structure that forbids these operators to high orders.

In contrast, the quality requirement for the massless quark solution is far less stringent. Beginning in the UV with a PQ symmetry at $\mathrm{Re}(M) = \mathrm{Im}(M) = 0$, instantons violate the PQ symmetry but not CP to produce an additive renormalization of M in the direction $\mathrm{Re}(M) > 0$. A PQ-violating operator with an $\mathcal{O}(1)$ phase results in an additional contribution to M in a general direction in the complex plane, but as long as the overall size of this contribution is small, then the resulting M is near the real axis and $\bar{\theta} \sim \mathrm{Im}(M)/\mathrm{Re}(M)$ stays small.

In the standard model massless up-quark solution, the leading dimensionful PQ-violating gauge-invariant operator is $|H|^2\tilde{H}Q\bar{u}/M_{Pl}^2$, and the effects of this operator must only compete with the dimensionless Yukawa couplings. An additional such "bare" contribution to the up-quark mass does not upset the massless up-quark solution so long as $\mathrm{Im}(m_u) \lesssim \bar{\theta} m_u$, which is easily satisfied as $v^2/M_{\rm Pl}^2 \sim 10^{-32}$.

Now, in comparison to the SM solution, which we know is not realized in nature, the model above necessarily needs new gauge dynamics at larger scales. Thus, there will be irrelevant operators with UV scalars that obtain larger vevs, for example, $|\Phi|^2\tilde{H}Q\bar{d}/M_{Pl}^2$ or $\tilde{H}Q\Sigma\bar{d}/M_{Pl}$. In the former case, these operators are completely safe so long as $\langle\Phi\rangle \sim \Lambda_9 \lesssim 10^{13}$ GeV, while in the latter case, there are no naturalness concerns so long as $\langle\Sigma\rangle \sim \Lambda_3 \lesssim 10^8$ GeV. Then, there are no issues with the model of Sec. III even if these PQ-violating operators are generated by quantum gravity with $O(1)$ coefficients and random phases.

Overall, rather than posing a problem, these naturalness considerations motivate the part of this model's parameter space in which the effects are most visible in low-energy experiments. In other words, while generally Λ_3 could be at some high scale and the model still works well, the above quantum-gravitational concerns suggest BSM quark flavor-changing physics should be, at worst, still in striking distance in the mid-to-long term [131].

IV. FURTHER ISSUES AND DIRECTIONS

In this work, we have uncovered and explored a new feature of colored standard model fermions. Remarkably, the known particle spectrum admits gauged flavor symmetries that bear out noninvertible symmetry structures through which ultraviolet instantons may resolve infrared naturalness issues. In the quark sector, this noninvertible symmetry appears when there is a nontrivial global structure for the gauge group—a possibility that is permitted because $N_c = N_g$. Following our infrared noninvertible symmetry analysis, we have sketched a scheme for implementing the massless down-type quark solution into the strong CP problem in $SU(9) \to SU(3)^2/\mathbb{Z}_3 \to SU(3)_C$. There are many directions for further investigation, and we briefly discuss some of them, in no particular order.

Alternative symmetry breaking.—It is potentially appealing to skip the $SU(3)_H$ phase of the theory entirely and go directly from $SU(9)$ to $(SU(3)_C \times U(1)_H)/\mathbb{Z}_3$ before breaking to $SU(3)_C$. From the UV, this route might allow us to avoid introducing the large Higgs representation 165, which is called for by $SU(9) \to SU(3)^2/\mathbb{Z}_3$, which would imply less suppression of the instanton density. From the IR, the Abelian horizontal symmetry is constrained only to a few TeV; see Refs. [132,133] for constraints from the LHC and low-energy flavor, and Ref. [134] for the general framework. An Abelian horizontal phase then allows potentially more visible observational signatures of such theories at the energy frontier.

Full flavor.—The model we studied above did not have enough structure to generate a realistic flavor sector, and we can easily understand the sense in which we are missing ingredients. Keeping the Higgs H as a color-flavor singlet means that the flavor-singlet coupling to the SM Higgs is allowed and generically leads to $y_u, y_d \propto \mathbb{1} +$ perturbations. Unfortunately, the observed SM quark spectrum is not so amenable to fitting this structure while keeping a good notion of power counting for the perturbations. Likely, the needed solution is to embed the Higgs in a color-flavor adjoint. The SM Higgs then arises as a linear combination of the colorless, flavor adjoint $80 \supset (1, 8)$, and the orientation of the Higgs in this matrix provides another handle for producing the SM Yukawa structure. This sort of structure has recently been utilized in Ref. [135] in a UV theory which is conceptually related to ours.

Interplay with other strong CP strategies.—We have utilized a mechanism of producing Hermitian Yukawas to ensure our UV boundary condition $\bar{\theta}(\Lambda_9)$ is not disrupted while generating the CKM matrix, but there may be other choices. The addition of vectorlike quarks would allow much greater freedom in shaping the Yukawas but may require a Nelson-Barr-like structure to avoid introducing dangerous phases. On the other hand, our avoidance of vectorlike quarks makes the scheme of Sec. III C seem well suited as a novel sort of model for spontaneous CP violation, which we will explore in the forthcoming work, Ref. [123]. There may also be a role for parity symmetry in a left-right extension of this model, where the broken $SU(2)_R$ is responsible for the initial distinction between up and down quarks. Another obvious strategy we are currently pursuing [136] is to spontaneously break the $U(1)_{PQ}$ to find a model for a heavy QCD axion with additional mass provided by the small instantons, whose phenomenology has been much discussed in recent years (see, e.g., Refs. [64,94,95,99,137–147]) and whose domain walls have rich phenomenology due to the noninvertible symmetry [22].

Relation to Agrawal and Howe [93].—The scenario of Agrawal and Howe also utilizes small instantons to revive the massless quark solution to the strong CP problem but not in a unified manner. They employ a flavor deconstruction of $SU(3)_C \subset SU(3)_1 \times SU(3)_2 \times SU(3)_3$ and begin with three free diagonal Yukawas (top, charm, and down); then, at the breaking scale Λ_3, the instantons of their three different gauge groups generate the three other diagonal Yukawas with appropriate sizes. As a result, the only model building they need to do is generate the off-diagonal entries. Indeed, by cleverly adding a few fields with judiciously chosen charges, they are able to fit the full CKM structure of the SM including δ_{CKM}. They then find that this process induces $\bar{\theta}$ as a two-loop correction that depends on the small off-diagonal Yukawas of the SM quarks and thus can be slightly below the current upper bound, $\Delta\bar{\theta} \lesssim 10^{-10}$, and within experimental reach.

We note that (at least structurally) one could attempt to embed the model of Agrawal and Howe in the same UV theory that we have, $SU(9) \supset SU(3)^3$. With the aesthetic and reductionist appeal of unification, one would have additional challenges of explaining how to obtain the appropriate flavor-asymmetric PQ symmetries, as well as the varying sizes of nonzero Yukawas and gauge couplings $g_1 > g_2 > g_3$, rather than inputting these by hand. A particular difficulty in this case may be that the up-type–down-type hierarchy flips in the first generation relative to the others, and it seems nontrivial for such a PQ to appear, starting with a flavor-unified theory. Still, it would be interesting to study in more detail a unified model embedding Agrawal and Howe, which would mirror the strategy of Davighi and Tooby-Smith [135], implementing flavor deconstruction and reunification in the electroweak sector. If such a scheme can work, the two-step breaking pattern from $SU(9)$ to $SU(3)_C$ has the advantage of generating the diagonal and off-diagonal structures at different scales but misses out on having the instantons and flavor breaking at different scales.

Quark-lepton color-flavor unification.—We were led to this model by a parallel with our earlier work on neutrino masses protected by noninvertible symmetries [3]. A unification of these parallel stories generating both tiny Dirac neutrino masses and small, strong CP violations may be achievable in a flavor-twisted Pati-Salam theory, $SU(12) \times SU(2)_L \times SU(2)_R$, which we are currently investigating. Such nontrivial gauge-flavor unified theories have received astonishingly little attention [135,148,149] and are ripe for further phenomenological study.

Very generally, the discovery of such unified theories gives interesting top-down motivation to study separate gauged quark and lepton flavor symmetries. It is evident from experimental bounds that the scale at which non-Abelian flavor physics is generated in the quark sector is far beyond the reach of the energy frontier. Naively, one might conclude that if we ultimately want quark-lepton unification, this should imply that there is a single flavor symmetry acting on both, meaning that the non-Abelian lepton flavor scale is also necessarily very high. This is not the case, as breaking $SU(12) \to SU(9)_{\mathrm{quarks}} \times SU(3)_{\mathrm{leptons}} \times U(1)_{B-L}$ means that lepton flavor may yet be generated at accessible energy scales consistently with unification. In particular, this idea provides further UV motivation for a muon collider to explore the energy frontier of lepton flavor physics (see, e.g., Ref. [150]).

Higgsed lattice field theory.—To fully understand the phenomenological potential of this model, we would like to know the bottom-to-top Yukawa ratio generated by the $SU(9)$ instantons over the entire parameter space. Since perturbative corrections are enhanced by N in large-N gauge theories for a given gauge coupling α, we quickly lose control, so numerical computations will be necessary to accurately understand the physics. Lattice computations of Higgs-Yang-Mills theories have received relatively little attention recently, but for early discussions, see, e.g., Refs. [151–153]. Here, we also have the challenge of trying to understand the size of the $SU(9)$ instanton effects without needing to specify a fully realistic theory in which we could check that we land on the SM spectrum of hadrons. One could perhaps input a potential which only condenses the scalar field Φ effecting $SU(9) \to SU(3)^2/\mathbb{Z}_3$, letting the flavor group confine as well, and inferring the ratio of yukawas from the relative masses of the flavor-symmetric $\Delta^{++} = uct$ and $\Delta^{-} = dsb$.

ACKNOWLEDGMENTS

We are grateful to T. Daniel Brennan, Richard Brower, Antonio Delgado, Samuel Homiller, Adam Martin, Kantaro Ohmori, and Lian-Tao Wang for useful conversations, and we thank Samuel Homiller and Adam Martin for comments on a draft of this work. Feynman diagrams have been created using FeynGame [154]. We thank the Aspen Center for Physics (supported by National Science Foundation Grant No. PHY-2210452) for the opportunity to participate in a summer workshop in 2023, during which part of this work was completed. The work of C. C. is supported by DOE Grant No. DE-SC0024367, by the Simons Collaboration on Global Categorical Symmetries, and by the Sloan Foundation. The work of S. H. is supported by the National Research Foundation of Korea (NRF) Grant No. RS-2023-00211732, by the Samsung Science and Technology Foundation under Project No. SSTF-BA2302-05, and by the POSCO Science Fellowship of POSCO TJ Park Foundation. The work of S. K. has been supported by an Oehme Postdoctoral Fellowship and an Arthur H. Compton Lectureship from the EFI at UChicago, and by NSF Grant No. PHY-2112540. S. K. additionally thanks the Simons Center for Geometry and Physics at Stony Brook University, the Mainz Institute for Theoretical Physics (MITP) of the Cluster of Excellence PRISMA+ (Project ID No. 39083149), and the Korea Advanced Institute of Science and Technology for hospitality during the completion of this work.

APPENDIX A: FRACTIONAL AND CFU INSTANTONS

In this appendix, we review a few facts about fractional instantons. Noninvertible symmetries appearing in this work are connected to a certain kind of fractional instanton configuration that may be loosely thought of as a special linear combination of fractional instantons of more than one gauge group factor, e.g., $SU(3)_C \times SU(3)_H$. They are examples of color-flavor-$U(1)$ (CFU) instantons introduced in Ref. [104] (see also Ref. [155]). The discussion presented here closely follows Refs. [32,104,156].

1. Fractional instantons

The simplest example of a gauge theory with fractional instantons is $PSU(N) = SU(N)/\mathbb{Z}_N$ gauge theory, and we will focus on it here. Let us first consider an $SU(N)$ gauge theory with only matter fields in the adjoint representation. In this case, the entire set of fields, including the gauge fields, are invariant under $\mathbb{Z}_N$ center transformations, which in turn means that there are $\mathbb{Z}_N$-charged Wilson lines that cannot be screened since there is no charged particle that can cut those lines: Such a theory enjoys 1-form electric $\mathbb{Z}_N^{(1)}$ symmetry. Since it is a good quantum symmetry, it (or any subgroup of it) can be gauged. If we gauge the entire $\mathbb{Z}_N^{(1)}$ electric center, the resulting theory is $SU(N)/\mathbb{Z}_N$ gauge theory. The latter no longer has 1-form electric symmetry, but now it has 1-form dual magnetic $\mathbb{Z}_N^{(1)}$ center symmetry. This fact can also be checked from the fact that $\pi_1(SU(N)/\mathbb{Z}_N) = \mathbb{Z}_N$. As we describe now, the path integral of $SU(N)/\mathbb{Z}_N$ gauge theory includes a summation over dynamical 2-form $\mathbb{Z}_N$ gauge fields B_2, which turn on $\mathbb{Z}_N$-valued magnetic flux,

$$\oint_{\Sigma_2} \frac{B_2}{2\pi} = \frac{1}{N}\mathbb{Z}. \tag{A1}$$

In addition, the theory contains fractional instanton configurations, as well as regular integer-valued instantons.

To see this case more explicitly, we start with $U(N) = \{[SU(N) \times U(1)]/\mathbb{Z}_N\}$ gauge theory and reduce it down to $SU(N)/\mathbb{Z}_N$ in two steps. First, we project down the local degree of freedom associated with the $U(1)$ factor by adding a Lagrange multiplier term to the action of the $U(N)$ theory,

$$S = \frac{1}{g^2}\int \mathrm{Tr}(\hat{f}_2 \wedge *\hat{f}_2) + \frac{i}{2\pi}\int F_2 \wedge \mathrm{Tr}(\hat{f}_2) + \frac{i\theta}{8\pi^2}\int \mathrm{Tr}(\hat{f}_2 \wedge \hat{f}_2). \tag{A2}$$

Here, $\hat{f}_2$ is the 2-form field strength of the $U(N)$ gauge field $\hat{a}_1$, and F_2 is a 2-form Lagrange multiplier whose equation of motion imposes $\mathrm{Tr}(\hat{f}_2) = 0$ which projects out the $U(1)$ degree of freedom. This theory is just $SU(N)$ and thus still has a $\mathbb{Z}$-valued instanton spectrum. To achieve $SU(N)/\mathbb{Z}_N$ theory with fractional instantons, we impose 1-form $\mathbb{Z}_N^{(1)}$ symmetry. To this end, we introduce the

$SU(N)$ gauge field a_1 and its field strength f_2. They are related to $U(N)$ fields as

$$\hat{a}_1 = a_1 + \frac{1}{N}A_1\mathbb{1}, \qquad \hat{f}_2 = f_2 + \frac{1}{N}dA_1\mathbb{1}, \qquad \text{(A3)}$$

where A_1 is a $U(1)$ gauge field and $\mathbb{1}$ is a $U(1)$ is the $N \times N$ identity matrix. Next, we impose $\mathbb{Z}_N^{(1)}$ symmetry under which fields transform as

$$\begin{aligned} &a_1 \to a_1, && A_1 \to A_1 - N\lambda_1 \\ \Rightarrow\ &\hat{a}_1 \to \hat{a}_1 - \lambda_1\mathbb{1}, && \hat{f}_2 \to \hat{f}_2 - d\lambda_1\mathbb{1}, \end{aligned} \qquad \text{(A4)}$$

where λ_1 is the 1-form symmetry transformation parameter. One then realizes that the action (A2) is not invariant under the above symmetry transformation. It can be made invariant by introducing a 2-form field B_2 and assigning a symmetry transformation rule $B_2 \to B_2 - d\lambda_1$. Then, the invariant action is given by

$$\begin{aligned} S = &\frac{1}{g^2}\int \mathrm{Tr}[(\hat{f}_2 - B_2\mathbb{1}) \wedge *(\hat{f}_2 - B_2\mathbb{1})] \\ &+ \frac{i}{2\pi}\int F_2 \wedge \mathrm{Tr}(\hat{f}_2 - B_2\mathbb{1}) \\ &+ \frac{i\theta}{8\pi^2}\int \mathrm{Tr}[(\hat{f}_2 - B_2\mathbb{1}) \wedge (\hat{f}_2 - B_2\mathbb{1})]. \end{aligned} \qquad \text{(A5)}$$

The appearance of the combination $(\hat{f}_2 - B_2\mathbb{1})$ can be thought of as coupling the theory to the 2-form gauge field B_2 of the $\mathbb{Z}_N^{(1)}$ electric symmetry. To understand the instanton spectrum, we first note that $\mathrm{Tr}(\hat{f}_2) = dA_1$. Then, the Lagrange multiplier term turns into

$$\frac{i}{2\pi}\int F_2 \wedge (dA_1 - NB_2), \qquad \text{(A6)}$$

which describes $\mathbb{Z}_N$ gauge theory (often called a BF theory). Specifically, the term $[(iN)/(2\pi)]F_2 \wedge B_2$ corresponds to the usual Lagrangian for the $\mathbb{Z}_N$ BF theory, and the above action is obtained by dualizing the 1-form gauge field in $F_2 = d\tilde{A}_1$ to A_1. One may interpret Eq. (A5) as $SU(N)$ theory coupled to this $\mathbb{Z}_N$ BF theory [157]. A brief review on BF theory can be found in Appendix B of Ref. [21], and for more detailed discussions, we refer to Refs. [32,157,158]. For instance, the equation of motion for F_2 sets $B_2 = dA_1/N$, showing that B_2 is a $\mathbb{Z}_N$ gauge field.

Finally, the θ-angle term is given by

$$\begin{aligned} S_\theta &= \frac{i\theta}{8\pi^2}\int \mathrm{Tr}(\hat{f}_2 \wedge \hat{f}_2) - NB_2 \wedge B_2 \\ &= \frac{i\theta}{8\pi^2}\int \mathrm{Tr}(\hat{f}_2 \wedge \hat{f}_2) - \mathrm{Tr}(\hat{f}_2) \wedge \mathrm{Tr}(\hat{f}_2) \\ &\quad + \frac{i\theta}{8\pi^2}\int N(N-1)B_2 \wedge B_2 \\ &= i\theta n + i\theta\left(\frac{N-1}{2N}\right)\int w_2 \wedge w_2. \end{aligned} \qquad \text{(A7)}$$

In the second line, we subtracted and added the combination $\mathrm{Tr}(\hat{f}_2) \wedge \mathrm{Tr}(\hat{f}_2)$ so that the first integral becomes the standard integer-valued $SU(N)$ instanton number n, while the second term captures the fractional instanton effects. To obtain the last line, we defined $w_2 = N[B_2/(2\pi)]$ (called the second Stiefel-Whitney class), whose integral is integer valued,

$$\oint w_2 = 0, 1, \ldots, (N-1). \qquad \text{(A8)}$$

In a theory where fermions can be introduced (technically, a spin structure can be defined), $\int w_2 \wedge w_2 \in 2\mathbb{Z}$, which clearly shows that the second term in the last line of Eq. (A7) indeed corresponds to the fractional instanton.

2. CFU instantons

Having described the fractional instantons of $SU(N)/\mathbb{Z}_N$ theory, we now briefly discuss CFU (color-(non-Abelian) flavor-$U(1)$) instantons [104]. The original construction given in Ref. [104] is the most general, and it refines background fields for the 1-form electric center symmetry of $SU(N)$ gauge theory. The novelty is that, given fermion contents with non-Abelian flavor symmetry—say, $SU(F)$ and a global $U(1)$ symmetry—CFU background analysis yields the most general electric 1-form symmetry of the theory. This then allows us to determine the largest set of mixed 0-form and 1-form 't Hooft anomalies, hence the strongest constraints on the IR phases of the strongly coupled gauge theory.

Let us denote the set of matter content as $\{\Psi_i\}$. In addition, let us write G_c, G_f, and G_u for color, non-Abelian flavor, and $U(1)$ groups. However, the following analysis can be generalized to any choice and any number of groups. Then, we write 0-form center transformations acting on Ψ_i's as $z_c \in Z(G_c), z_f \in Z(G_f)$, and $z_u \in Z(G_u)$. Calling Wilson lines of G_c, G_f, and G_u as W_c, W_f, and W_u, the 1-form electric symmetry can be figured out by understanding the spectrum of topologically protected (i.e., unscreened by local charges) Wilson lines, including all possible composite Wilson lines of the form $W_c^\ell W_f^m W_u^n, \ell, m, n \in \mathbb{Z}$. The protected Wilson line spectrum is obtained by solving so-called "cocycle conditions." The basic idea is to determine the most general set of center transformations that act trivially on the *entire* $\{\Psi_i\}$ (and gauge fields),

$$\Psi_i \to z_c z_f z_u \Psi_i = \Psi_i \quad \forall\, i. \tag{A9}$$

Since the fields of the theory are uncharged under any such center transformations, associated (composite) Wilson lines are not screened; therefore, electric 1-form symmetry is determined. Focusing on the case $G_c = SU(N)$, $G_f = SU(F)$, and $G_u = U(1)$, the solutions take the form

$$z_c = e^{\frac{2\pi i}{N}\ell}, \qquad z_f = e^{\frac{2\pi i}{F}m}, \qquad z_u = e^{2\pi i n}, \tag{A10}$$

with $\ell = 0, \ldots, (N-1), m = 0, \ldots, (F-1), n \in [0,1]$. Then, we can show that a particular combination (determined by $\{\ell, m, n\}$) of 2-form background gauge fields—called B_c, B_f, B_u—of individual 1-form electric center symmetry can be consistently activated. If no solution with only a single center factor exists, e.g., $(\ell \neq 0, m = 0, n = 0)$, 1-form electric symmetry is only defined in terms of composite Wilson operators; in that case, one has to turn on a specific combination of 2-form background fields controlled by the solution $\{\ell, m, n\}$ as an acceptable configuration.

Nonvanishing 2-form background gauge fields lead to CFU instantons (or CFU topological charge). Using the expression of fractional instanton, Eq. (A7), it is given by the combination of the following three (C, F, and U) fractional instantons:

$$\mathcal{N}_c = \left(\frac{N-1}{N}\right)\int \frac{w_c \wedge w_c}{2} = \ell_1 \ell_2 \left(1 - \frac{1}{N}\right), \tag{A11}$$

$$\mathcal{N}_f = \left(\frac{F-1}{F}\right)\int \frac{w_f \wedge w_f}{2} = m_1 m_2 \left(1 - \frac{1}{F}\right), \tag{A12}$$

$$\mathcal{N}_u = \int \frac{B_u \wedge B_u}{8\pi^2} = n_1 n_2, \tag{A13}$$

where w fields are defined in terms of B fields as in the previous section [see below Eq. (A7)]. Also, when $\{\ell_1, m_1, n_1\} \neq \{\ell_2, m_2, n_2\}$, the above expression means the following. Consider the spacetime manifold of the form $M_4 = \mathbb{S}^2 \times \mathbb{S}^2$. Then, given two solutions $\{\ell_1, m_1, n_1\}$ and $\{\ell_2, m_2, n_2\}$ to the cocycle conditions, we take a background configuration that is a formal sum of $\{\ell_1, m_1, n_1\}$ units of CFU fluxes piercing the first $\mathbb{S}^2$ with zero flux through the second $\mathbb{S}^2$ and a configuration with zero flux through the first $\mathbb{S}^2$ and $\{\ell_2, m_2, n_2\}$ units of CFU fluxes through the second $\mathbb{S}^2$. For the color part, this process leads to

$$\oint_{\mathbb{S}^2 \times \mathbb{S}^2} \frac{w_c \wedge w_c}{2} = \ell_1 \ell_2, \tag{A14}$$

and similarly for the flavor and $U(1)$ parts. Finally, we stress again that $\mathcal{N}_c$, $\mathcal{N}_f$, and $\mathcal{N}_u$ separately are not well-defined configurations. Only the whole combination makes sense and gives rise to integer-valued (i.e., sensible) Dirac indices. When the electric 1-form symmetries are gauged, then the background fields $B_{c,f,u}$ are path-integrated dynamical fields, and the CFU configurations turn into dynamical instantons of the theory. See Ref. [104] for more details and for usage of CFU instantons.

APPENDIX B: GLOBAL STRUCTURE AND NONINVERTIBLE SYMMETRIES

1. SM global structure

In Sec. II, we learned that the noninvertible symmetries of $(SU(3)_C \times U(1)_H)/\mathbb{Z}_3$ or $(SU(3)_C \times SU(3)_H)/\mathbb{Z}_3$ depend sensitively on the global structure, and the effects of interest are not present in the absence of this "modding" of the gauge group. With the goal of providing additional background to readers unfamiliar with these concepts, in this appendix, we review the global structure of the SM itself, $G_{\mathrm{SM}} = (SU(3)_C \times SU(2)_L \times U(1)_Y)/\Gamma$, and conclude that the SM noninvertible symmetries do not depend on the choice $\Gamma \in \{\mathbb{1}, \mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_6\}$. We refer to Ref. [159] for a discussion of the line operator spectrum (Wilson, 't Hooft, and dyonic) depending on the SM global structure, to Ref. [12] for the (fractional) instanton spectrum and related cosmological impacts, and to Ref. [22] for the discussion of SM global structure and axion noninvertible symmetry, as well as the resulting observational implications in terms of axion domain-wall physics.

In Table I, we show the representations of the SM fermions under the gauge symmetries and classical global symmetries of the SM, in addition to the right-handed neutrinos and the Higgs boson. One may observe that there are certain combinations of center transformations that act trivially on all of the fields of the SM. For example, the center of $SU(2)_L$ is a $\mathbb{Z}_2$ subgroup whose nontrivial element acts as $(-1)\mathbb{1}_2$ on the fields charged under $SU(2)_L$. These effects can be compensated exactly by a rotation by π of hypercharge, the $\mathbb{Z}_2 \subset U(1)_Y$ subgroup; thus, if we perform a diagonal transformation by $(-\mathbb{1}_2)e^{i\pi Y}$, then, e.g.,

$$Q_i \to Q_i(-1)e^{i\pi(+1)} = Q_i, \qquad L_i \to L_i(-1)e^{i\pi(-3)} = L_i,$$
$$\bar{u}_i \to \bar{u}_i e^{i\pi(-4)} = \bar{u}_i, \ldots, \tag{B1}$$

and all the SM fields are invariant under such a rotation [recall that the W gauge fields are in the adjoint, two-index representation of $SU(2)$, so they transform as $(-1)^2$]. The fact is that the observed fields with odd numbers of $SU(2)$ indices also have odd hypercharge, while those with even numbers of $SU(2)$ indices have even hypercharge. It is not difficult to see that the quarks have just the right hypercharges that a similar diagonal $\mathbb{Z}_3$ subgroup of $SU(3)_C \times U(1)_Y$ also acts trivially on the known fields.

In fact, it is not too difficult to show that the entire SM fields, matter as well as gauge fields, are invariant under $\mathbb{Z}_6$ center transformations. To show this explicitly, we first note that the centers of $SU(3)_C$ and $SU(2)_L$ and the $\mathbb{Z}_6$ center of $U(1)_Y$ are generated by

$$e^{2\pi i\lambda_8/3} = e^{2\pi i/3}\mathbb{1}_3 \in SU(3)_C,$$
$$e^{2\pi i T_3/2} = -\mathbb{1}_2 \in SU(2)_L, \qquad e^{2\pi i q_Y/6} \in U(1)_Y, \tag{B2}$$

where $\lambda_8 = \mathrm{diag}(1,1,-2)$ and $T_3 = \mathrm{diag}(1,-1)$, and q_Y denotes the generator of $U(1)_Y$ normalized as $e^{2\pi i q_Y} = 1$. Then, the statement is that the representations of SM fields are such that the combination

$$e^{2\pi i\lambda_8/3} \times e^{2\pi i T_3/2} \times e^{2\pi i q_Y/6} \tag{B3}$$

acts trivially (i.e., as an identity) on all fields of the SM. Realizing Eq. (B3) as the generator of $\mathbb{Z}_6$ proves that the whole SM is uncharged under $\mathbb{Z}_6$.

In terms of 1-form symmetry and related global structure, this result can be explained as follows. Let us denote a general Wilson line of the SM as $W_C^a W_L^b W_Y^c$, where W_C, W_L, and W_Y are charge-1 Wilson lines of $SU(3)_C$, $SU(2)_L$, and $U(1)_Y$, respectively, and $a, b, c \in \mathbb{Z}$. This particular Wilson line can be thought of as the worldline of a probe particle whose representation is such that it carries N-ality (roughly the number of boxes of Young tableau) of (a, b, c) under the Lie algebra $su(3)_C \times su(2)_L \times u(1)_Y$. If there exists a dynamical, light particle with N-ality that is a divisor of (a, b, c), then pair production can cut the Wilson line, and we say that the corresponding line is screened. In the absence of dynamical charges to screen, on the other hand, the Wilson line is stable, and we can discuss 1-form global symmetry. In particular, such a topologically protected Wilson line shows the existence of electric 1-form center symmetry. Representations of SM are then consistent with the existence of $\mathbb{Z}_6$ stable Wilson lines and $\mathbb{Z}_6^{(1)}$ 1-form electric symmetry, which is the case when the global structure is $\Gamma = \mathbb{1}$.

Nontrivial $\Gamma = \mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_6$ is obtained if electric $\mathbb{Z}_6^{(1)}$ or its subgroup is gauged. Thus, the 2-form background gauge field of the 1-form electric center, denoted before as B_2, is now path integrated. If we gauge $\mathbb{Z}_q \subset \mathbb{Z}_6$, $q = 2, 3, 6$, we obtain $\Gamma = \mathbb{Z}_q$. Gauging an electric symmetry generates the dual magnetic symmetry [160]. Indeed, with $\mathbb{Z}_q \subset \mathbb{Z}_6$ gauged, the SM now has $\mathbb{Z}_{6/q}^{(1)}(\mathrm{e}) \times \mathbb{Z}_q^{(1)}(\mathrm{m})$, where the first (second) is the electric (magnetic) 1-form symmetry.

2. Absence of noninvertible symmetries of the SM

In this appendix, we show that, with the up Yukawas y_u turned on, the SM does not possess any noninvertible symmetries acting on quark fields for all choices of global structure Γ.

To this end, we recall that noninvertible symmetry can exist if there is global $U(1)$ that is primarily broken by $U(1)$ or fractional instantons. In this sense, it is already clear from Table II that noninvertible symmetry is not present in the SM with any choice of $\Gamma \in \{\mathbb{Z}_6, \mathbb{Z}_3, \mathbb{Z}_2\}$: ABJ anomaly coefficients are flavor universal, and crucially, the $SU(3)_C$ instanton already achieves maximal breaking of all $U(1)$ symmetries of the theory. While fractional instantons, in general, lead to smaller anomaly coefficients, and hence more breaking of $U(1)$ symmetry, the above fact clearly shows that the best $\mathbb{Z}_6$ fractional instantons of $SU(3)_C \times SU(2)_L \times U(1)_Y/\mathbb{Z}_6$ can do is to break $U(1)$ symmetries as much as regular $SU(3)_C$ instantons do. What is left is to demonstrate that this is indeed the case.

A $\mathbb{Z}_6$ fractional instanton configuration can be constructed following Appendix A. The $\mathbb{Z}_6$ fractional instanton is given by a combination of the $\mathbb{Z}_3$ instanton of $SU(3)_C/\mathbb{Z}_3$, the $\mathbb{Z}_2$ instanton of $SU(2)_L/\mathbb{Z}_2$, and the instanton of $U(1)_Y/\mathbb{Z}_6$. Following the notation defined in Appendix A, we can write topological charges as

$$\mathcal{N}_C = \ell_1\ell_2\left(1 - \frac{1}{N_c}\right), \qquad \mathcal{N}_L = m_1 m_2\left(1 - \frac{1}{2}\right),$$
$$\mathcal{N}_Y = n_1 n_2. \tag{B4}$$

A consistent, minimal $\mathbb{Z}_6$ fractional instanton can be found by solving cocycle conditions [104]. The result is the combination of $\mathcal{N}_C$, $\mathcal{N}_L$, and $\mathcal{N}_Y$ with $\ell_1 = \ell_2 = -1$, $m_1 = m_2 = 1$, and $n_1 = n_2 = -\frac{1}{6}$. Calling this configuration [CLY], topological charges take values $\mathcal{N}_C = \frac{2}{3}$, $\mathcal{N}_L = \frac{1}{2}$, and $\mathcal{N}_Y = \frac{1}{36}$. The ABJ anomalies of $U(1)_{\tilde{B}_i}$ and $U(1)_{\bar{d}_i}$ are independent of the flavor index i and are computed as

$$\tilde{B}_i[\mathrm{CLY}] = 1, \qquad \bar{d}_i[\mathrm{CLY}] = 1. \tag{B5}$$

We conclude that $\mathbb{Z}_6$ fractional instantons break $U(1)$ symmetries $\Pi_i U(1)_{\tilde{B}_i} \times U(1)_{\bar{d}_i}$ of the SM with $y_d = 0$ just as much as non-Abelian $SU(3)_C$ instantons do; Therefore, no noninvertible symmetry exists with any choice of Γ.

APPENDIX C: 't HOOFT VERTICES IN $SU(9)$

In the quantum theory of $SU(9)$, nonperturbative gauge-theoretic effects generate 't Hooft vertices that dynamically violate the anomalous quantum numbers. As depicted in Fig. 2, this process results in the generation of the down-type Yukawa, which, in a general basis, is parametrically

$$\mathcal{L} \sim y_t^\star e^{i\theta_9} e^{-S_{\mathrm{inst}}} H\mathbf{Q}\bar{\mathbf{d}}, \tag{C1}$$

where, with $S_{\rm inst} = \{(8\pi^2)/[g_9^2(\Lambda_9)]\}$, we obtain the leading estimate for the instanton effects. Note that the instanton-induced down-type Yukawa is $y_b \propto y_t^\star$, which ensures that the effective $\bar{\theta}$ still vanishes exactly: Using chiral rotations, one may move phases in y_t and y_b into the θ term, and they exactly cancel. This finding is just an explicit manifestation of the observation made in Ref. [87].

These 't Hooft vertices are precisely analogous to those that generate Dirac neutrino masses in the lepton $SU(3)_H$ theory of Ref. [3]. However, our purpose here is to generate the bottom Yukawa—which is not parametrically small —$y_b/y_t \sim 1/40$, which requires a more precise understanding of the instanton-induced interaction. In particular, we must include the polynomial factor that arises from quadratic fluctuations around the instanton background. We will ultimately find that, in the parameter space where large y_b could be generated, the one-instanton computation is not fully trustworthy, and the honest viability of this scenario will need to be determined by lattice simulations. Thus, we aim at demonstrating the plausibility of generating y_b of the right order of magnitude within this one-instanton approximation, and we discuss the features of the quadratic fluctuations that enter the result only at this level.

The one-loop instanton gas computation was first performed in the pioneering work Ref. [161] by 't Hooft, and for our usage, it essentially consists of calculating

$$\langle H(x_1)\mathbf{Q}(x_2)\bar{\mathbf{d}}(x_3)\rangle_{\text{1-instanton}}, \tag{C2}$$

which is the three-point correlator summed over all possible 1-instanton backgrounds, and then performing Lehmann-Symanzik-Zimmermann reduction to obtain the three-point amplitude. As effective field theorists, we can think of finding the Wilsonian action obtained by integrating out instantons at smaller sizes ρ (ρ is the size modulus of the instanton) by matching to the amplitude $\mathcal{M}(H, \mathbf{Q}, \bar{\mathbf{d}})$. Eventually, we will integrate out "all" the instanton effects at the scale $\rho \sim \Lambda_9^{-1}$. To justify this procedure in a Higgsed gauge theory, recall that, in the case of a 1-instanton background, the action for the gauge field at the scale ρ^{-1} is given by

$$e^{-8\pi^2/g^2(\rho)} = e^{-8\pi^2/g^2(\Lambda_{\rm UV})}(\rho\Lambda_{\rm UV})^{b_0}, \tag{C3}$$

with $\Lambda_{\rm UV}$ the UV cutoff scale and $b_0 > 0$ the 1-loop β-function coefficient. However, the contribution from the scalar that achieves spontaneous breaking of the gauge group at Λ_9 behaves as

$$e^{-8\pi^2\rho^2\Lambda_9^2}, \tag{C4}$$

and we see that large instantons are exponentially suppressed [98,99,162]. On the other hand, from Eq. (C3), we see that the gauge contribution becomes more important for larger ρ since the gauge coupling increases. In fact, the balance between the two contributions shows that the dominant contribution comes from $\rho^2 \approx b_0/16\pi^2\Lambda_9^2$, i.e., roughly $\rho \sim 1/\Lambda_9$, as can be studied systematically in the formalism of "constrained instantons" [162].

The correlator we want to calculate can be computed by a path integral

$$\int \mathcal{D}A\mathcal{D}\phi_i\mathcal{D}\psi_i H\mathbf{Q}\bar{\mathbf{d}}e^{-S_{\rm gauge}-\int\mathcal{L}_{\rm int}}, \tag{C5}$$

integrating over the gauge field A in the sector $\int F\tilde{F}/(32\pi^2) = 1$ and also any charged scalars ϕ or fermions ψ. For a detailed accounting of such integrals, we refer to, for example, Refs. [99,163–166], and we give solely an overview here. There are a number of effects that appear in this path integration, starting with the evaluation of bosonic modes.

(i) *Instanton background action*—The semiclassical gauge field background contributes to the action as

$$\begin{aligned} S &= \int \frac{1}{4g^2}F^2 \\ &= \int \frac{1}{8g^2}(F \pm \tilde{F})^2 \mp \int \frac{1}{4g^2}F\tilde{F}, \end{aligned} \tag{C6}$$

having made use of $F^2 = \tilde{F}^2$ and completing the square. Now, we see that in a nontrivial topological sector with $\int F\tilde{F}/(32\pi^2) = k$, $k \in \mathbb{Z}$, the action has a minimum at $S_{1-\rm inst} = 8\pi^2|k|/g^2$ for (anti)self-dual field configurations $F = \pm\tilde{F}$. Thus, the effects of the nontrivial saddles are suppressed by $\exp(-2\pi/\alpha)$, with $\alpha \equiv g^2/4\pi$.

In a general basis where the UV theory also contains a θ term, $\mathcal{L} \supset [(i\theta)/(32\pi^2)]F\tilde{F}$, this phase enters the correlation functions in the instanton background by simply evaluating this boundary term on the background solution. This process provides the θ angle in Eq. (C1) in exactly the right combination to preserve $\bar{\theta} = 0$.

(ii) *Gauge field zero modes*—The classical 1-instanton background, using the 't Hooft symbols $\eta_{a\mu\nu}$ and written in the "singular gauge,"

$$A_\mu(x) = \frac{2\rho^2}{(x-x_0)^2}\frac{\eta_{a\mu\nu}(x-x_0)^\nu J^a}{(x-x_0)^2 + \rho^2}, \tag{C7}$$

possesses four translational zero modes corresponding to x_0, and $4N - 5$ orientational zero modes corresponding to the $SU(2) \subset SU(N)$ subgroup generators J^a. The final modulus is the size ρ, which is quantum mechanically nontrivial because the coupling evolves with scale. These zero modes are dealt with by the method of collective coordinates, and the proper normalization of these integrals leads to factors of $2^{-2N}(2\pi/\alpha(\rho))^{2N}$. The integration over gauge rotations yields a further factor proportional to $1/(N-1)!(N-2)!$ [167].

(iii) *Charged field nonzero modes*—At the level of quadratic fluctuations about the instanton background, the matter field energy levels are skewed slightly away from their vacuum values. Note that 't Hooft showed that the determinants take the same form regardless of the statistics of the field, and the result depends solely on the degrees of freedom in different representations of the subgroup in which the instanton resides. In total, there is a factor $\exp[-(-1)^F \sum a(t)n(t)]$, where $a(t)$ is a certain polynomial 't Hooft derived that depends on the "isospin" t of a given representation [taking the values $a(0)=0$, $a(\frac{1}{2})=0.146$, $a(1)=0.443$, and $a(\frac{3}{2})=0.853$], $n(t)$ counts the number of such representations, and fermions contribute inversely because of Grassmann statistics. The adjoint gauge field contributes $n(1)=1$ and $n(\frac{1}{2})=2(N-2)$, while the SM fundamental matter counts as $(-1)^F n(\frac{1}{2})=-4$.

Then, after having integrated over all the bosonic modes except for the scaling modulus (and also the nonzero fermionic modes), the instanton density in an $SU(N)$ gauge theory takes the form

$$D(\rho)=\frac{4}{\pi^2}\frac{2^{-2N}e^{-\sum(-1)^F a(t)n(t)}}{(N-1)!(N-2)!}\left(\frac{2\pi}{\alpha(\rho)}\right)^{2N}\exp\left(-\frac{2\pi}{\alpha(\rho)}\right), \tag{C8}$$

where ρ is the size parameter of the instanton solution.

Charged fermion zero modes.—Now, for charged fermions, there is a qualitatively new sort of effect. In an instanton background, charged fermions will have zero modes that are exactly counted by the Dirac index appearing in the coefficient of chiral anomalies. The path integral includes integrals over these zero modes, which, for our content of solely fundamental fermions, implies one zero-mode integral per field,

$$\int \mathcal{D}\psi_i \supset \int \prod_{i=1}^{N_f} d\xi_i^{(0)}, \tag{C9}$$

where $\xi_i^{(0)}$ is the Grassmann coefficient of the zero-mode eigenspinor of the ith field. For a Grassmann variable ξ, integration works as $\int d\xi = 0$ and $\int d\xi\xi = 1$, so the presence of these Grassmann zero-mode integrals implies that correlation functions in the instanton background all vanish unless they include fermion fields to be integrated against those zero modes. Thus, with just the SM fields, the 1-instanton path integral immediately generates a 't Hooft vertex that contributes to correlation functions like

$$\langle \mathbf{Q}\bar{\mathbf{d}}\mathbf{Q}\bar{\mathbf{u}}\rangle. \tag{C10}$$

Instead, to reach the Yukawa operator of interest, we need an insertion of a Lagrangian operator, meaning that the nonvanishing contribution appears in the path integral as

$$\int \mathcal{D}(\text{fields})H(x_1)\mathbf{Q}(x_2)\bar{\mathbf{d}}(x_3) \times\left(\int d^4y\, i y_t^\star \tilde{H}^\dagger(y)\mathbf{Q}^\dagger(y)\bar{\mathbf{u}}^\dagger(y)\right)e^{-\int\mathcal{L}}, \tag{C11}$$

corresponding to the vertex insertion depicted in Fig. 2. Despite what may look like loops of fermionic lines in the figure, there is no undetermined loop momentum there to integrate over—rather, we integrate against the zero-mode wave function that is provided by the 't Hooft vertex, which does not result in any loop suppression.

Past the SM matter and the $SU(9)$ gauge field, the charged scalar fields we introduce to effect symmetry breaking modify the 't Hooft vertex by contributing to the instanton density. In Sec. III B, the breaking $SU(9)\to SU(3)^2/\mathbb{Z}_3$ was achieved by the vev of a three-index symmetric tensor $\Phi^{\{ABC\}}$, which decomposes under an $SU(2)$ subgroup to contribute $n(\frac{3}{2})=1$, $n(1)=7$, $n(\frac{1}{2})=28$, and $n(0)=84$. In Sec. III C, our toy model of flavor breaking utilized $N_s=2$ scalar adjoints $\Sigma^A_{1,2}$, which contribute the same way as the vector, and a single scalar fundamental, which contributes oppositely to one of the quarks.

Then, we end up with a bottom Yukawa interaction, after the extra fermion modes have been taken care of, which we can write in the form of an integral over the instanton scale with an infrared cutoff Λ_9^{-1},

$$\mathcal{L}\supset y_t^\star e^{i\theta}\int_0^{\Lambda_9^{-1}}\frac{d\rho}{\rho}D(\rho)H\mathbf{Q}\bar{\mathbf{d}}. \tag{C12}$$

This interaction is sensitive to both the beta function at scales around Λ_9 and the spectrum of charged matter.

Finally, to evaluate this scale integral, we must account for the nontrivial evolution of the gauge coupling as a function of the scale. We factor out the prefactor coming from the integration of quadratic fluctuations of charged fields, which does not depend on the scale. At one loop, the α evolution is $2\pi/\alpha(\rho)=2\pi/\alpha(\Lambda_9)-\beta_9\log(\rho\Lambda_9)$, leading to an integral over the size modulus, which eventually looks like

$$\int\frac{d\rho}{\rho}(\rho\Lambda_9)^{\beta_9}\left(\frac{2\pi}{\alpha(\Lambda_9)}-\beta_9\log(\rho\Lambda_9)\right)^{2N}. \tag{C13}$$

The prefactor in the instanton density Eq. (C8) is extremely small for large N, so it is unclear if large y_b can be

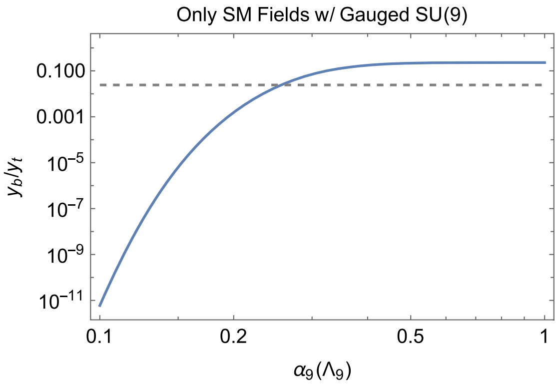

FIG. 6. Size of y_b/y_t generated by $SU(9)$ instantons at Λ_9, performing the integral in Eq. (C12) using the instanton density of Eq. (C8) in the theory of solely the SM matter fields. This theory is not realistic because it does not yet incorporate matter involved in breaking $SU(9) \to SU(3)_C$. The dashed line is the observed infrared value of y_b/y_t.

generated. However, the integral can be quite large, which is evident for large α as the integral approaches $(2N)!/\beta_9$; thus, it can potentially compete with the prefactor. Overall, for large N, $(2N)!/(N-1)!(N-2)! \sim 2^{2N}\sqrt{N^5/\pi}$, which indeed compensates for the small constant factors. The result is found in Fig. 6.

[1] P. A. M. Dirac, *New basis for cosmology*, Proc. R. Soc. A **165**, 199 (1938).

[2] G. 't Hooft, *Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking*, NATO Sci. Ser. B **59**, 135 (1980).

[3] C. Córdova, S. Hong, S. Koren, and K. Ohmori, *Neutrino masses from generalized symmetry breaking*, Phys. Rev. X **14**, 031033 (2024).

[4] Our analysis also revealed a way to generate Majorana neutrino masses by nonperturbative quantum effects.

[5] C. Córdova, T. T. Dumitrescu, and K. Intriligator, *Exploring 2-group global symmetries*, J. High Energy Phys. 02 (2019) 184.

[6] Z. Wan and J. Wang, *Beyond Standard Models and grand unifications: Anomalies, topological terms, and dynamical constraints via cobordisms*, J. High Energy Phys. 07 (2020) 062.

[7] J. Davighi, B. Gripaios, and N. Lohitsiri, *Global anomalies in the Standard Model(s) and beyond*, J. High Energy Phys. 07 (2020) 232.

[8] J. Wang, *Anomaly and cobordism constraints beyond the Standard Model: Topological force*, arXiv:2006.16996.

[9] T. D. Brennan and C. Córdova, *Axions, higher-groups, and emergent symmetry*, J. High Energy Phys. 02 (2022) 145.

[10] Y. Hidaka, M. Nitta, and R. Yokokura, *Global 3-group symmetry and 't Hooft anomalies in axion electrodynamics*, J. High Energy Phys. 01 (2021) 173.

[11] J. Fan, K. Fraser, M. Reece, and J. Stout, *Axion mass from magnetic monopole loops*, Phys. Rev. Lett. **127**, 131602 (2021).

[12] M. M. Anber and E. Poppitz, *Nonperturbative effects in the Standard Model with gauged 1-form symmetry*, J. High Energy Phys. 12 (2021) 055.

[13] J. Wang, Z. Wan, and Y.-Z. You, *Cobordism and deformation class of the standard model*, Phys. Rev. D **106**, L041701 (2022).

[14] J. Wang and Y.-Z. You, *Gauge enhanced quantum criticality between grand unifications: Categorical higher symmetry retraction*, arXiv:2111.10369.

[15] J. Wang and Y.-Z. You, *Gauge enhanced quantum criticality beyond the standard model*, Phys. Rev. D **106**, 025013 (2022).

[16] J. McNamara and M. Reece, *Reflections on parity breaking*, arXiv:2212.00039.

[17] C. Córdova and S. Koren, *Higher flavor symmetries in the Standard Model*, Ann. Phys. (N.Y.) **535**, 2300031 (2023).

[18] Y. Choi, H. T. Lam, and S.-H. Shao, *Non-invertible time-reversal symmetry*, Phys. Rev. Lett. **130**, 131602 (2023).

[19] J. Wang, Z. Wan, and Y.-Z. You, *Proton stability: From the standard model to beyond grand unification*, Phys. Rev. D **106**, 025016 (2022).

[20] R. Yokokura, *Non-invertible symmetries in axion electrodynamics*, arXiv:2212.05001.

[21] T. D. Brennan, S. Hong, and L.-T. Wang, *Coupling a cosmic string to a TQFT*, J. High Energy Phys. 03 (2024) 145.

[22] C. Córdova, S. Hong, and L.-T. Wang, *Axion domain walls, small instantons, and non-invertible symmetry breaking*, arXiv:2309.05636.

[23] Y. Choi, M. Forslund, H. T. Lam, and S.-H. Shao, *Quantization of axion-gauge couplings and non-invertible higher symmetries*, Phys. Rev. Lett. **132**, 121601 (2024).

[24] M. van Beest, P. Boyle Smith, D. Delmastro, Z. Komargodski, and D. Tong, *Monopoles, scattering, and generalized symmetries*, arXiv:2306.07318.

[25] T. D. Brennan, *A new solution to the Callan Rubakov effect*, arXiv:2309.00680.

[26] Y. Choi, H. T. Lam, and S.-H. Shao, *Non-invertible Gauss law and axions*, J. High Energy Phys. 09 (2023) 067.

[27] C. Córdova and K. Ohmori, *Quantum duality in electromagnetism and the fine-structure constant*, arXiv:2307.12927.

[28] M. Reece, *Axion-gauge coupling quantization with a twist*, J. High Energy Phys. 10 (2023) 116.

[29] P. Putrov and J. Wang, *Categorical symmetry of the Standard Model from gravitational anomaly*, Phys. Rev. D **110**, 125028 (2024).

[30] M. van Beest, P. Boyle Smith, D. Delmastro, R. Mouland, and D. Tong, *Fermion-monopole scattering in the Standard Model*, arXiv:2312.17746.

[31] D. Aloni, E. García-Valdecasas, M. Reece, and M. Suzuki, *Spontaneously broken (−1)-form U(1) symmetries*, arXiv:2402.00117.

[32] T. D. Brennan and S. Hong, *Introduction to generalized global symmetries in QFT and particle physics*, arXiv:2306.00912.

[33] L. Bhardwaj, L. E. Bottini, L. Fraser-Taliente, L. Gladden, D. S. W. Gould, A. Platschorre *et al.*, *Lectures on generalized symmetries*, Phys. Rep. **1051**, 1 (2024).

[34] P. R. S. Gomes, *An introduction to higher-form symmetries*, SciPost Phys. Lect. Notes **74**, 1 (2023).

[35] R. Luo, Q.-R. Wang, and Y.-N. Wang, *Lecture notes on generalized symmetries and applications*, arXiv:2307.09215, 10.1016/j.physrep.2024.02.002.

[36] S.-H. Shao, *What's done cannot be undone: TASI lectures on non-invertible symmetry*, arXiv:2308.00747.

[37] S. Schafer-Nameki, *ICTP lectures on (non-)invertible generalized symmetries*, Phys. Rep. **1063**, 1 (2024).

[38] B. C. Allanach, B. Gripaios, and J. Tooby-Smith, *Semisimple extensions of the Standard Model gauge algebra*, Phys. Rev. D **104**, 035035 (2021).

[39] S. Dar, *The neutron EDM in the SM: A review*, arXiv:hep-ph/0008248.

[40] C. Abel *et al.*, *Measurement of the permanent electric dipole moment of the neutron*, Phys. Rev. Lett. **124**, 081803 (2020).

[41] Throughout this work, we will abuse terminology in referring to $\tilde{J}$ as the "Jarlskog invariant." In fact, it differs from the original definition by Jarlskog [42] by a real proportionality factor. Explicitly, Jarlskog defined $iC \equiv [y_u^\dagger y_u, y_d^\dagger y_d]$ and showed that

$$\det C = -2FF'J,$$
$$F = (m_t - m_c)(m_t - m_u)(m_c - m_u)/m_t^3,$$
$$F' = (m_b - m_s)(m_b - m_d)(m_s - m_d)/m_b^3,$$
$$J = \mathrm{Im}\,(V_{ij}V_{kl}V_{kj}^*V_{il}^*) \propto \sin\delta_{\mathrm{CKM}},$$

where V is the CKM matrix and δ_{CKM} is the CP-odd phase. While J sensibly splits off the observed dependence on the mass eigenvalues and has been measured as $J \approx 3 \times 10^{-5}$ [43], $\tilde{J} \equiv \det C$ can be written simply in terms of the Yukawas and thus is an easier diagnostic of nonvanishing CP violation.

[42] C. Jarlskog, *Commutator of the quark mass matrices in the standard electroweak model and a measure of maximal CP nonconservation*, Phys. Rev. Lett. **55**, 1039 (1985).

[43] R. L. Workman *et al.* (Particle Data Group), *Review of particle physics*, Prog. Theor. Exp. Phys. **2022**, 083C01 (2022).

[44] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the 2π decay of the K_2^0 meson*, Phys. Rev. Lett. **13**, 138 (1964).

[45] J. R. Ellis and M. K. Gaillard, *Strong and weak CP violation*, Nucl. Phys. **B150**, 141 (1979).

[46] A. E. Nelson, *Naturally weak CP violation*, Phys. Lett. B **136**, 387 (1984).

[47] S. M. Barr, *Solving the strong CP problem without the Peccei-Quinn symmetry*, Phys. Rev. Lett. **53**, 329 (1984).

[48] L. Bento, G. C. Branco, and P. A. Parada, *A minimal model with natural suppression of strong CP violation*, Phys. Lett. B **267**, 95 (1991).

[49] S. M. Barr, D. Chang, and G. Senjanovic, *Strong CP problem and parity*, Phys. Rev. Lett. **67**, 2765 (1991).

[50] K. S. Babu and R. N. Mohapatra, *CP violation in seesaw models of quark masses*, Phys. Rev. Lett. **62**, 1079 (1989).

[51] K. S. Babu and R. N. Mohapatra, *A solution to the strong CP problem without an axion*, Phys. Rev. D **41**, 1286 (1990).

[52] N. Craig, I. Garcia Garcia, G. Koszegi, and A. McCune, *P not PQ*, J. High Energy Phys. 09 (2021) 130.

[53] M. Dine and P. Draper, *Challenges for the Nelson-Barr mechanism*, J. High Energy Phys. 08 (2015) 132.

[54] A. Valenti and L. Vecchi, *The CKM phase and $\bar{\theta}$ in Nelson-Barr models*, J. High Energy Phys. 07 (2021) 203.

[55] L. Vecchi, *Spontaneous CP violation and the strong CP problem*, J. High Energy Phys. 04 (2017) 149.

[56] A. Valenti and L. Vecchi, *Perturbative running of the topological angles*, J. High Energy Phys. 01 (2023) 131.

[57] J. de Vries, P. Draper, and H. H. Patel, *Do minimal parity solutions to the strong CP problem work?*, arXiv:2109.01630.

[58] P. Asadi, S. Homiller, Q. Lu, and M. Reece, *Chiral Nelson-Barr models: Quality and cosmology*, Phys. Rev. D **107**, 115012 (2023).

[59] T. Banks, M. Dine, and N. Seiberg, *Irrational axions as a solution of the strong CP problem in an eternal universe*, Phys. Lett. B **273**, 105 (1991).

[60] R. Kuchimanchi, *Solution to the strong CP problem: Supersymmetry with parity*, Phys. Rev. Lett. **76**, 3486 (1996).

[61] R. N. Mohapatra and A. Rasin, *Simple supersymmetric solution to the strong CP problem*, Phys. Rev. Lett. **76**, 3490 (1996).

[62] G. Hiller and M. Schmaltz, *Solving the strong CP problem with supersymmetry*, Phys. Lett. B **514**, 263 (2001).

[63] G. Hiller and M. Schmaltz, *Strong weak CP hierarchy from nonrenormalization theorems*, Phys. Rev. D **65**, 096009 (2002).

[64] A. Hook, *Anomalous solutions to the strong CP problem*, Phys. Rev. Lett. **114**, 141801 (2015).

[65] A. Albaid, M. Dine, and P. Draper, *Strong CP and* SUZ_2, J. High Energy Phys. 12 (**2015**) 046.

[66] N. Kaloper and J. Terning, *Landscaping the strong CP problem*, J. High Energy Phys. 03 (2019) 032.

[67] M. Carena, D. Liu, J. Liu, N. R. Shah, C. E. M. Wagner, and X.-P. Wang, *ν solution to the strong CP problem*, Phys. Rev. D **100**, 094018 (2019).

[68] A. L. Cherchiglia and C. C. Nishi, *Solving the strong CP problem with non-conventional CP*, J. High Energy Phys. 03 (2019) 040.

[69] D. Dunsky, L. J. Hall, and K. Harigaya, *Higgs parity, strong CP, and dark matter*, J. High Energy Phys. 07 (2019) 016.

[70] A. Valenti and L. Vecchi, *Super-soft CP violation*, J. High Energy Phys. 07 (2021) 152.

[71] J. Wang, *Strong CP problem and symmetric mass solution*, arXiv:2212.14036.

[72] G. Dvali, *Strong-CP with and without gravity*, arXiv:2209.14219.

[73] H.-Y. Cheng, *The strong CP problem revisited*, Phys. Rep. **158**, 1 (1988).

[74] M. Dine, *TASI lectures on the strong CP problem*, in Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 2000): Flavor physics for the millennium (2000), pp. 349–369 [arXiv:hep-ph/0011376].

[75] A. Hook, *TASI lectures on the strong CP problem and axions*, Proc. Sci., TASI2018 (**2019**) 004 [arXiv:1812.02669].

[76] M. Reece, *TASI lectures: (No) Global symmetries to axion physics*, Proc. Sci., TASI2022 (**2024**) 008 [arXiv:2304.08512].

[77] We will often use the language of differential forms to emphasize when quantities are topological, and we remind the reader that $\mathrm{Tr}(G\wedge G)=\frac{1}{4}G^{\mu\nu}\tilde{G}_{\mu\nu}$, with $\tilde{G}^{\mu\nu}=\frac{1}{2}\epsilon^{\mu\nu\rho\sigma}G_{\rho\sigma}$.

[78] S. L. Adler, *Axial vector vertex in spinor electrodynamics*, Phys. Rev. **177**, 2426 (1969).

[79] J. S. Bell and R. Jackiw, *A PCAC puzzle: $\pi^0\to\gamma\gamma$ in the σ model*, Nuovo Cim. A **60**, 47 (1969).

[80] R. D. Peccei and H. R. Quinn, *CP conservation in the presence of instantons*, Phys. Rev. Lett. **38**, 1440 (1977).

[81] R. D. Peccei and H. R. Quinn, *Constraints imposed by CP conservation in the presence of instantons*, Phys. Rev. D **16**, 1791 (1977).

[82] C. Vafa and E. Witten, *Parity conservation in QCD*, Phys. Rev. Lett. **53**, 535 (1984).

[83] S. Weinberg, *A new light boson?*, Phys. Rev. Lett. **40**, 223 (1978).

[84] F. Wilczek, *Problem of strong P and T invariance in the presence of instantons*, Phys. Rev. Lett. **40**, 279 (1978).

[85] H. M. Georgi and I. N. McArthur, *Instantons and the u quark mass*, Harvard University, Lyman Lab. Phys. Technical Report No. HUTP-81/A011, 1981.

[86] D. B. Kaplan and A. V. Manohar, *Current mass ratios of the light quarks*, Phys. Rev. Lett. **56**, 2004 (1986).

[87] K. Choi, C. W. Kim, and W. K. Sze, *Mass renormalization by instantons and the strong CP problem*, Phys. Rev. Lett. **61**, 794 (1988).

[88] M. Srednicki, *Comment on "Ambiguities in the up-quark mass"*, Phys. Rev. Lett. **95**, 059101 (2005).

[89] T. Banks, Y. Nir, and N. Seiberg, *Missing (up) mass, accidental anomalous symmetries, and the strong CP problem, in 2nd IFT Workshop on Yukawa Couplings and the Origins of Mass*, pp. 26–41 [arXiv:hep-ph/9403203].

[90] M. Dine, P. Draper, and G. Festuccia, *Instanton effects in three flavor QCD*, Phys. Rev. D **92**, 054004 (2015).

[91] S. Aoki *et al.* (Flavour Lattice Averaging Group), *FLAG review 2019*, Eur. Phys. J. C **80**, 113 (2020).

[92] C. Alexandrou, J. Finkenrath, L. Funcke, K. Jansen, B. Kostrzewa, F. Pittler *et al.*, *Ruling out the massless up-quark solution to the strong **CP** problem by computing the topological mass contribution with lattice QCD*, Phys. Rev. Lett. **125**, 232001 (2020).

[93] P. Agrawal and K. Howe, *A flavorful factoring of the strong CP problem*, J. High Energy Phys. 12 (2018) 035.

[94] P. Agrawal and K. Howe, *Factoring the strong CP problem*, J. High Energy Phys. 12 (2018) 029.

[95] B. Holdom and M. E. Peskin, *Raising the axion mass*, Nucl. Phys. **B208**, 397 (1982).

[96] J. M. Flynn and L. Randall, *A computation of the small instanton contribution to the axion potential*, Nucl. Phys. **B293**, 731 (1987).

[97] K. Choi and H. D. Kim, *Small instanton contribution to the axion potential in supersymmetric models*, Phys. Rev. D **59**, 072001 (1999).

[98] C. Csaki and H. Murayama, *Instantons in partially broken gauge groups*, Nucl. Phys. **B532**, 498 (1998).

[99] C. Csáki, M. Ruhdorfer, and Y. Shirman, *UV sensitivity of the axion mass from instantons in partially broken gauge groups*, J. High Energy Phys. 04 (2020) 031.

[100] R. S. Gupta, V. V. Khoze, and M. Spannowsky, *Small instantons and the strong CP problem in composite Higgs models*, Phys. Rev. D **104**, 075011 (2021).

[101] Y. Choi, H. T. Lam, and S.-H. Shao, *Noninvertible global symmetries in the Standard Model*, Phys. Rev. Lett. **129**, 161601 (2022).

[102] C. Córdova and K. Ohmori, *Non-invertible chiral symmetry and exponential hierarchies*, Phys. Rev. X **13**, 011034 (2023).

[103] The technique employed in the appendixes follows that of Ref. [104], which uses equations called "cocycle conditions" to systematically determine the most general possible global structure of the gauge group and spectrum of fractional instantons, including those utilizing Abelian factors of the gauge group. These configurations are called "color-flavor-$U(1)$" (CFU) instantons in Ref. [104].

[104] M. M. Anber, S. Hong, and M. Son, *New anomalies, TQFTs, and confinement in bosonic chiral gauge theories*, J. High Energy Phys. 02 (2022) 062.

[105] D. Gaiotto, A. Kapustin, N. Seiberg, and B. Willett, *Generalized global symmetries*, J. High Energy Phys. 02 (2015) 172.

[106] Y. Choi, C. Córdova, P.-S. Hsin, H. T. Lam, and S.-H. Shao, *Noninvertible duality defects in $3+1$ dimensions*, Phys. Rev. D **105**, 125016 (2022).

[107] J. Kaidi, K. Ohmori, and Y. Zheng, *Kramers-Wannier-like duality defects in $(3+1)$D gauge theories*, Phys. Rev. Lett. **128**, 111601 (2022).

[108] K. Roumpedakis, S. Seifnashri, and S.-H. Shao, *Higher gauging and non-invertible condensation defects*, arXiv:2204.02407, 10.1007/s00220-023-04706-9.

[109] C. Córdova, K. Ohmori, and T. Rudelius, *Generalized symmetry breaking scales and weak gravity conjectures*, J. High Energy Phys. 11 (**2022**) 154.

[110] We also note that this theory has an (approximate) higher group structure $SU(2)^{(0)}_{Q,u,d}\times_\kappa U(1)^{(1)}_H$, but it will not play an important role for our purposes (cf. Ref. [17]).

[111] For completeness, we also note that the Q_1d_2 and Q_2d_1 components of the down-Yukawa matrix are further protected by the two anomaly-free $U(1)$ symmetries.

[112] We note that χ has the same charge assignment as a right-handed neutrino, though it need not be one. However, if it is, this model introduces no new fermions past the $16N_g$ of a standard model.

[113] We may just as well shift $\tilde{B}+\bar{d}$ by any good symmetry and use such a new direction as the Peccei-Quinn symmetry, for instance, $\bar{d}$.

[114] Axion theories, which also make use of good $U(1)_{PQ}$ symmetries, famously have a "quality problem," with quantum gravitational violations of this global symmetry posing severe fine-tuning issues. Despite also requiring a good PQ symmetry, our quality requirements are far less stringent, so there is no naturalness issue here, as we discuss in Sec. III E.

[115] Y. Grossman and P. Tanedo, *Just a taste: Lectures on flavor physics, in Theoretical Advanced Study Institute in Elementary Particle Physics: Anticipating the next*

discoveries in particle physics (2018), pp. 109–295 [arXiv:1711.03624],

[116] K. A. Intriligator and N. Seiberg, *Duality, monopoles, dyons, confinement and oblique confinement in supersymmetric* $\mathrm{SO}(n_c)$ *gauge theories*, Nucl. Phys. **B444**, 125 (1995).

[117] R. N. Mohapatra and G. Senjanovic, *natural suppression of strong P and T non-invariance*, Phys. Lett. B **79**, 283 (1978).

[118] J. L. Evans, B. Feldstein, W. Klemm, H. Murayama, and T. T. Yanagida, *Hermitian flavor violation*, Phys. Lett. B **703**, 599 (2011).

[119] Of course, the generation of $\Lambda_3 \ll \Lambda_9$ would constitute a hierarchy problem. However, note that, unlike for the electroweak hierarchy problem, the classical natural solutions applied to this hierarchy problem still work fine. Perhaps the real role of supersymmetry or warped extra dimensions in our Universe is to stabilize the scale at which flavor is generated, Λ_3.

[120] Vectorlike quarks would be the obvious way to address this issue, but they can be disastrous for our strong *CP* solution—they easily provide new sources of *CP* violation that are directly communicated to the SM quarks and, in any case, make our 't Hooft vertices more complicated and potentially lead to further suppression. Thus, we forego these quarks and utilize a strategy of communicating flavor-breaking purely bosonically.

[121] To match onto the perturbative QCD coupling, having started with large $SU(9)$ instanton effects needed to generate y_b/y_t, we assume all the $SU(9)$ partners of Σ and Φ obtain masses at the scale Λ_3 rather than Λ_9. Thus, we can benefit from the rapid RG running of gauge coupling from the strong to the weak regime as the theory flows from the UV $SU(9)$ phase to $SU(3)_C$ in the IR. We discuss this point in detail in Sec. III D.

[122] H. E. Haber and Z. Surujon, *A group-theoretic condition for spontaneous CP violation*, Phys. Rev. D **86**, 075007 (2012).

[123] S. Hong and S. Koren, *Flavorful spontaneous CP violation to solve the strong CP problem* (unpublished).

[124] G. Isidori, Y. Nir, and G. Perez, *Flavor physics constraints for physics beyond the Standard Model*, Annu. Rev. Nucl. Part. Sci. **60**, 355 (2010).

[125] J. Ellis, M. Madigan, K. Mimasu, V. Sanz, and T. You, *Top, Higgs, diboson and electroweak fit to the Standard Model effective field theory*, J. High Energy Phys. 04 (2021) 279.

[126] S. Ghigna, M. Lusignoli, and M. Roncadelli, *Instability of the invisible axion*, Phys. Lett. B **283**, 278 (1992).

[127] S. M. Barr and D. Seckel, *Planck scale corrections to axion models*, Phys. Rev. D **46**, 539 (1992).

[128] R. Holman, S. D. H. Hsu, T. W. Kephart, E. W. Kolb, R. Watkins, and L. M. Widrow, *Solutions to the strong CP problem in a world with gravity*, Phys. Lett. B **282**, 132 (1992).

[129] M. Kamionkowski and J. March-Russell, *Planck scale physics and the Peccei-Quinn mechanism*, Phys. Lett. B **282**, 137 (1992).

[130] P. A. Zyla *et al.* (Particle Data Group), *Review of particle physics*, Prog. Theor. Exp. Phys. (**2020**) 083C01.

[131] R. K. Ellis *et al.*, *Physics Briefing Book: Input for the European Strategy for Particle Physics Update 2020*, arXiv:1910.11775.

[132] J.-Y. Liu, Y. Tang, and Y.-L. Wu, *Searching for Z' gauge boson in an anomaly-free* $\mathrm{U}(1)'$ *gauge family model*, J. Phys. G **39**, 055003 (2012).

[133] A. Crivellin, G. D'Ambrosio, and J. Heeck, *Addressing the LHC flavor anomalies with horizontal gauge symmetries*, Phys. Rev. D **91**, 075006 (2015).

[134] P. Langacker and M. Plumacher, *Flavor changing effects in theories with a heavy Z' boson with family nonuniversal couplings*, Phys. Rev. D **62**, 013006 (2000).

[135] J. Davighi and J. Tooby-Smith, *Electroweak flavour unification*, J. High Energy Phys. 09 (2022) 193.

[136] C. Córdova, S. Hong, S. Koren, and L.-T. Wang, *The heavy QCD axion from color-flavor unification* (unpublished).

[137] B. Holdom, *Strong QCD at high-energies and a heavy axion*, Phys. Lett. B **154**, 316 (1985).

[138] V. A. Rubakov, *Grand unification and heavy axion*, JETP Lett. **65**, 621 (1997).

[139] Z. Berezhiani, L. Gianfagna, and M. Giannotti, *Strong CP problem and mirror world: The Weinberg-Wilczek axion revisited*, Phys. Lett. B **500**, 286 (2001).

[140] H. Fukuda, K. Harigaya, M. Ibe, and T. T. Yanagida, *Model of visible QCD axion*, Phys. Rev. D **92**, 015021 (2015).

[141] S. Dimopoulos, A. Hook, J. Huang, and G. Marques-Tavares, *A collider observable QCD axion*, J. High Energy Phys. 11 (2016) 052.

[142] M. K. Gaillard, M. B. Gavela, R. Houtz, P. Quilez, and R. Del Rey, *Color unified dynamical axion*, Eur. Phys. J. C **78**, 972 (2018).

[143] T. Gherghetta, N. Nagata, and M. Shifman, *A visible QCD axion from an enlarged color group*, Phys. Rev. D **93**, 115010 (2016).

[144] T. Gherghetta, V. V. Khoze, A. Pomarol, and Y. Shirman, *The axion mass from 5D small instantons*, J. High Energy Phys. 03 (2020) 063.

[145] A. Kivel, J. Laux, and F. Yu, *Supersizing axions with small size instantons*, J. High Energy Phys. 11 (2022) 088.

[146] A. Valenti, L. Vecchi, and L.-X. Xu, *Grand color axion*, J. High Energy Phys. 10 (**2022**) 025.

[147] A. Hook, S. Kumar, Z. Liu, and R. Sundrum, *High quality QCD axion and the LHC*, Phys. Rev. Lett. **124**, 221801 (2020).

[148] T.-K. Kuo and N. Nakagawa, *The generation problem and the symplectic group*, Nucl. Phys. **B250**, 641 (1985).

[149] J. Davighi, A. Greljo, and A. E. Thomsen, *Leptoquarks with exactly stable protons*, Phys. Lett. B **833**, 137310 (2022).

[150] A. Dasgupta, P. S. B. Dev, T. Han, R. Padhan, S. Wang, and K. Xie, *Searching for heavy leptophilic Z': From lepton colliders to gravitational waves*, J. High Energy Phys. 12 (2023) 011.

[151] R. C. Brower, D. A. Kessler, T. Schalk, H. Levine, and M. Nauenberg, *The* $\mathrm{SU}(2)$ *adjoint Higgs model*, Phys. Rev. D **25**, 3319 (1982).

[152] I.-H. Lee and J. Shigemitsu, *Spectrum calculations in the lattice Georgi-Glashow model*, Nucl. Phys. **B263**, 280 (1986).

[153] R. E. Shrock, *Lattice Higgs models*, Nucl. Phys. B, Proc. Suppl. **4**, 373 (1988).

[154] R. V. Harlander, S. Y. Klein, and M. Lipp, *FeynGame*, Comput. Phys. Commun. **256**, 107465 (2020).

[155] M. M. Anber and E. Poppitz, *On the baryon-color-flavor (BCF) anomaly in vector-like theories*, J. High Energy Phys. 11 (2019) 063.

[156] For more reviews on generalized symmetries, see Refs. [33–36] and the appendixes of Ref. [21].

[157] A. Kapustin and N. Seiberg, *Coupling a QFT to a TQFT and duality*, J. High Energy Phys. 04 (2014) 001.

[158] T. Banks and N. Seiberg, *Symmetries and strings in field theory and gravity*, Phys. Rev. D **83**, 084019 (2011).

[159] D. Tong, *Line operators in the Standard Model*, J. High Energy Phys. 07 (2017) 104.

[160] Y. Tachikawa, *On gauging finite subgroups*, SciPost Phys. **8**, 015 (2020).

[161] G. 't Hooft, *Computation of the quantum effects due to a four-dimensional pseudoparticle*, Phys. Rev. D **14**, 3432 (1976).

[162] I. Affleck, *On constrained instantons*, Nucl. Phys. **B191**, 429 (1981).

[163] S. Vandoren and P. van Nieuwenhuizen, *Lectures on instantons*, arXiv:0802.1862.

[164] D. E. Morrissey, T. M. P. Tait, and C. E. M. Wagner, *Proton lifetime and baryon number violating signatures at the CERN LHC in gauge extended models*, Phys. Rev. D **72**, 095003 (2005).

[165] A. Ringwald, *High-energy breakdown of perturbation theory in the electroweak instanton sector*, Nucl. Phys. **B330**, 1 (1990).

[166] O. Espinosa, *High-energy behavior of baryon and lepton number violating scattering amplitudes and breakdown of unitarity in the Standard Model*, Nucl. Phys. **B343**, 310 (1990).

[167] C. W. Bernard, *Gauge zero modes, instanton determinants, and QCD calculations*, Phys. Rev. D **19**, 3013 (1979).