Gauge Symmetry and the Theta-Vacuum

Richard Healey
rhealey@email.Arizona.edu

1 Two Kinds of Symmetry

Abstractly, a symmetry of a structure is an automorphism—a transformation that maps the elements of an object back onto themselves so as to preserve the structure of that object.

A physical theory specifies a set of models—mathematical structures—that may be used to represent various different situations, actual as well as merely possible, and to make claims about them. Any application of a physical theory is to a situation involving some system, actual or merely possible. Only rarely is that system the entire universe: typically, one applies a theory to some subsystem, regarded as a relatively isolated part of its world. The application proceeds by using the theory to model the situation of that subsystem in a way that abstracts from and idealizes the subsystem's own features, and also neglects or idealizes its interactions with the rest of the world.

We can therefore enquire about the symmetries of the class of models of a theory; or we can enquire about the symmetries of a class of situations, whether or not we have in mind a theory intended to model them. The first enquiry may reveal some *theoretical symmetry*: the second may reveal some *empirical symmetry*. An empirical symmetry can be recognized even without a physical theory to account for it. But it does not cease to be empirical if and when such a theory becomes available. A theory may entail an empirical symmetry.

Galileo illustrated his relativity principle by describing a famous empirical symmetry of this kind.

> Shut yourself up with some friend in the main cabin below decks on some large ship, and have with you there some flies, butterflies and other small flying animals... When you have observed all these things carefully..., have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that. You will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still.

His implicit claim is that a situation inside the cabin when the ship is in motion is indistinguishable from another situation inside the cabin when the ship is at rest by observations confined to those situations. The claim follows from a principle of the relativity of all uniform horizontal motion. While we know today that an unqualified form of Galileo's claim is false, in a modified form it continues to play an important role in physics.

Galileo's implicit claim is that situations related by a uniform collective horizontal motion are empirically symmetrical. Specifically

> A 1-1 mapping $\varphi : \mathcal{S} \rightarrow \mathcal{S}$ of a set of situations onto itself is an *empirical symmetry* if and only if any two situations related by φ are indistinguishable by means of measurements confined to each situation.

A measurement is confined to a situation just in case it is a measurement of intrinsic properties of (one or more objects in) that situation. Note that the reference to measurement is not superfluous here, in so far as a situation may feature unmeasurable intrinsic properties. If every function $\varphi \in \Phi$ is an empirical symmetry of S, then S is symmetric under Φ-transformations. Note that situations in S related by a transformation φ may be in the same or different possible worlds: if φ is an empirical symmetry, then $\varphi(s)$ may be in the same world w as s, but only if w is itself sufficiently symmetric.

One may distinguish symmetries of the set of situations to which a theory may be applied from symmetries of the set of the theory's models.

> A mapping $f : \mathcal{M} \rightarrow \mathcal{M}$ of the set of models of a theory Θ onto itself is a *theoretical symmetry of* Θ if and only if the following condition obtains: For every model m of Θ that may be used to represent (a situation s in) a possible world w, $f(m)$ may also be used to represent (s in) w.

If every function $f \in F$ is a symmetry of Θ, then Θ is symmetric under F-transformations. Theoretical symmetries may be purely formal features of a theory, in so far as they relate different but equivalent ways the theory has of representing one and the same empirical situation. One model may be more conveniently applied to a given situation than another model related to it by a theoretical symmetry, but the theoretical as well as empirical content of any claim made about that situation will be the same no matter which model is applied. But a theoretical symmetry of a theory may entail an empirical symmetry, in which case it is not a purely formal feature of the theory.

The empirical symmetry associated with uniform velocity boosts in special relativity is a consequence of a theoretical symmetry of special relativity, if one associates each model of that theory with an inertial frame with respect to which a given situation is represented. For then the strong empirical symmetry becomes a consequence of the Lorentz invariance of the theory—the fact that the Lorentz transform of any model is also a model of the theory. The Lorentz transform of any model may be used to represent the same situation as the original model (from the perspective of a boosted inertial frame); but it may also be used to represent a boosted *duplicate* of that situation (from the perspective of the original frame). (Here a duplicate of a situation is a situation that shares all its intrinsic properties.) The special theory of relativity entails the empirical symmetry associated with Lorentz invariance by implying that these empirically equivalent situations are not merely empirically indistinguishable by means of measurements confined to those situations, but indistinguishable by reference to any intrinsic properties or relations of entities each involves.

Relativity principles assert empirical symmetries. If "local" gauge transformations reflect some similar empirical symmetry, then they also represent distinct but indistinguishable situations. But I shall defend the conventional wisdom that the successful employment of Yang-Mills theories warrants the conclusion that "local" gauge transformations are only theoretical symmetries of these theories that reflect no corresponding non-trivial empirical symmetries among the situations they represent. "Local" gauge symmetry is a purely formal feature of these theories.

2 Warm-up Exercise: Faraday's Cube

Michael Faraday constructed a hollow cube with sides 12 feet long, covered it with good conducting materials but insulated it carefully from the ground, and electrified it so that it was at a large potential difference from the rest of his laboratory. He made the following entry in his diary in 1836:

"I went into this cube and lived in it, but though I used lighted candles, electrometers, and all other tests of electrical states, I could not find the least influence on them"

Both Faraday and Galileo described observations of symmetries in nature. In each case, different situations are compared, and it is noted that these are indistinguishable with respect to a whole class of phenomena. But while velocity boosts are paradigm empirical symmetries, gauge symmetry is usually taken to be a purely formal feature of a theory. In this case, adding the same constant to all electrical potentials is a symmetry of classical electromagnetism. Why doesn't Faraday's cube provide a perfect analogue of Galileo's ship for local gauge symmetry? (Note that the electric potential transformation $\varphi \to \varphi + a$ is an example of a local gauge transformation $A_\mu \to A_\mu + \partial_\mu \Lambda$ with $A_\mu = (\varphi, -\mathbf{A})$ and $\Lambda = at$.)

There is an important disanalogy between the Lorentz boost symmetry imperfectly illustrated by Galileo's ship and the local gauge symmetry illustrated by Faraday's cube. While both are theoretical symmetries of the relevant theories, only in the former case does this theoretical symmetry imply a corresponding empirical symmetry.

In order that charging the exterior of Galileo's cube should provide an example of the relevant kind of empirical symmetry, two conditions must be satisfied. It must produce a situation inside the cube that differs from its situation when uncharged in a way that corresponds to performing a local gauge transformation on its interior. But despite this difference, the transformed situation must remain internally indistinguishable from the original situation.

To see how it might be possible to meet both conditions, consider the analogous case of the Lorentz-boosted (space!)ship. Even though the situation inside the ship is a perfect duplicate of its situation before boosting, the theory itself implies that these situations are related by a boost transformation: because the only theoretical models that represent *both situations at once* are models in which the two situations are related by a velocity boost.

But classical electromagnetic theory has no analogous implication in the case of Galileo's cube. It contains models, each of which represents the cube both before and after charging, that represent the cube's interior as being in exactly the same state, independent of the charge on its exterior! There is no theoretical or experimental reason to suppose that charging the cube's exterior does anything to alter the electromagnetic state of its interior. Charging the exterior of Faraday's cube is not a way of performing a local gauge transformation on its interior: it is no more effective than painting it blue, or simply waiting for a day!

3 The $\theta-$Vacuum

The ground state of a quantized non-Abelian Yang-Mills gauge theory is usually described by a real-valued parameter θ—a fundamental new constant of nature. The structure of this vacuum state is often said to arise from a degeneracy of the vacuum of the corresponding classical theory. The degeneracy allegedly follows from the fact that "large" (but not "small") local gauge transformations connect physically distinct states of zero field energy. In a classical non-Abelian Yang-Mills gauge theory, "large" gauge transformations connect models of distinct but indistinguishable situations. This seems to show that at least "large" local gauge symmetry is an empirical symmetry.

In clarifying the distinction between "large" and "small" gauge transformations we will be driven to a deeper analysis of the significance of gauge symmetry. But understanding the θ-vacuum will require refining, not abandoning, the thesis that "local" gauge symmetry is a purely theoretical symmetry.

Before moving to the quantum theory, consider a classical SU(2) Yang-Mills gauge theory with action

$$S = \frac{1}{2g^2} \int Tr(\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu})d^4x \tag{1}$$

where $\mathbf{F}_{\mu\nu} = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu + [\mathbf{A}_\mu, \mathbf{A}_\nu]$

and $\mathbf{A}_\mu = A^j_\mu \frac{\sigma_j}{2i}$ transform as

$$\mathbf{A}_\mu \to \mathbf{A}'_\mu = \mathbf{U}\mathbf{A}_\mu\mathbf{U}^\dagger + (\partial_\mu \mathbf{U})\mathbf{U}^\dagger, \ \mathbf{F}_{\mu\nu} \to \mathbf{U}\mathbf{F}_{\mu\nu}\mathbf{U}^\dagger \tag{2}$$

under a "local" gauge transformation $\mathbf{U}(\mathbf{x}, t)$. (Here σ_j $(j = 1, 2, 3)$ are Pauli spin matrices.)

The field energy is zero if $\mathbf{F}_{\mu\nu} = 0$: that condition is consistent with $\mathbf{A}_\mu = 0$ and gauge transforms of this. Now restrict attention to those gauge transformations for which $\mathbf{A}'_0 = 0$, $\partial_0 \mathbf{A}'_j = 0$ i.e.

$$\mathbf{A}_\mu = 0 \to \mathbf{A}'_j(\mathbf{x}) = \{\partial_j \mathbf{U}(\mathbf{x})\}\mathbf{U}^\dagger(\mathbf{x}), \ \mathbf{A}'_0 = 0 \tag{3}$$

These are generated by functions $\mathbf{U} : \mathbb{R}^3 \to SU(2)$.

Those functions that satisfy $\mathbf{U}(\mathbf{x}) \to \mathbf{1}$ for $|\mathbf{x}| \to \infty$ constitute smooth maps $\mathbf{U} : S^3 \to SU(2)$, where S^3 is the 3-sphere. Some of these may be continuously deformed into the identity map $\mathbf{U}(\mathbf{x}) = \mathbf{1}$. But others cannot be. The maps divide into a countable set of equivalence classes, each characterized by an element of the homotopy group $\pi_3(SU(2)) = \mathbb{Z}$ called the *winding number*.

Maps in the same equivalence class as the identity map are said to generate "small" "local" gauge transformations: these are taken to relate alternative representations of the same classical vacuum. But $\mathbf{A}'_\mu$, $\mathbf{A}''_\mu$ generated from $\mathbf{A}_\mu = 0$ by maps $\mathbf{U}(\mathbf{x})$ from different equivalence classes are often said to represent *distinct* classical vacua, and $\mathbf{A}'_\mu$, $\mathbf{A}''_\mu$ are said to be related by "large" gauge transformations. (It is important to distinguish this claim from the quite different proposition, according to which degenerate *quantum* vacua may be related by a *"global"* gauge transformation in cases of spontaneous symmetry breaking. We are concerned at this point with a possible degeneracy in the *classical* vacuum of a non-Abelian Yang-Mills gauge theory.)

But if "local" gauge symmetry is a purely formal feature of a theory, then a gauge transformation cannot connect representations of physically distinct situations, even if it is "large"! And yet, textbook discussions of the *quantum* θ-vacuum typically represent this by a superposition of states, each element of which is said to correspond to a distinct state from the degenerate classical vacuum.

4 Two Analogies

Such discussions frequently appeal to a simple analogy from elementary quantum mechanics. Consider a particle moving in a one-dimensional periodic potential of finite height, like a sine wave. Classically, the lowest energy state is infinitely degenerate: the particle just sits at the bottom of one or other of the identical wells in the potential. But quantum mechanics permits tunnelling between neighboring wells, which removes the degeneracy. In the absence of tunnelling, there would be a countably infinite set of degenerate ground states of the form $\psi_n(x) = \psi_0(x - na)$ where a is the period of the potential. These are related by the translation operator $\hat{T}_a$: $\hat{T}_a\psi(x) = \psi(x - a)$. $\hat{T}_a$ is unitary and commutes with the Hamiltonian $\hat{H}$. Hence there are joint eigenstates $|\theta\rangle$ of $\hat{H}$ and $\hat{T}_a$ satisfying $\hat{T}_a|\theta\rangle = \exp(i\theta)|\theta\rangle$.

Such a state has the form

$$|\theta\rangle = \sum_{n=-\infty}^{+\infty} \exp\{-in\theta\} |n\rangle \tag{4}$$

where $\psi_n(x)$ is the wave function of state $|n\rangle$. When tunnelling is allowed for, the energy of these states depends on the parameter $\theta \in [0, 2\pi)$. It is as if quantum tunneling between the distinct classical ground states has removed the degeneracy, resulting in a spectrum of states of different energies parametrized by θ, each corresponding to a different superposition of classical ground states.

An alternative analogy is provided by a charged pendulum swinging from a long, thin solenoid whose flux Φ is generating a static Aharonov-Bohm potential $\mathbf{A}$. The Hamiltonian is

$$\hat{H} = \frac{1}{2m}[-i(\boldsymbol{\nabla} - ie\mathbf{A})]^2 + V \tag{5}$$

With a natural "tangential" choice of gauge for $\mathbf{A}$ this becomes

$$\hat{H} = -\frac{1}{2ml^2}\left(\frac{d}{d\omega} - ielA\right)^2 + V(\omega) \tag{6}$$

where the pendulum has mass m, charge e, length l and angle coordinate ω. If the wave function is transformed according to

$$\psi(\omega) = \exp\left[ie\int_0^\omega lAd\omega'\right]\varphi(\omega) \tag{7}$$

then the transformed wave function satisfies the Schrödinger equation with simplified Hamiltonian

$$\hat{H}_\varphi = -\frac{1}{2m}\frac{d^2}{d\omega^2} + V(\omega) \tag{8}$$

The boundary condition $\psi(\omega + 2\pi) = \psi(\omega)$ now becomes

$$\varphi(\omega + 2\pi) = \exp\{-ie\Phi\}\varphi(\omega) \tag{9}$$

which is of the same form as in the first analogy: $\hat{T}_{2\pi}\varphi = \exp\{i\theta\}\varphi$, with $\theta = -e\Phi$.

Unlike the periodic potential, the charged pendulum features a *unique* classical ground state. The potential barrier that would have to be overcome to "flip" the pendulum over its support can be tunnelled through quantum mechanically, but the tunnel ends up back where it started from! This produces a θ-dependent ground state energy as in the analogy of the periodic potential. But in this case there is a *single* state corresponding to an *external* parameter θ rather than a spectrum of states labeled by an internal parameter θ.

Which is the better analogy? Is the θ-vacuum in a quantized non-Abelian gauge theory more like a quantum state of the periodic potential, or a state of the charged quantum pendulum?

In his book *Classical Theory of Gauge Fields*, Rubakov describes both analogies. He notes that vacua of a classical Yang-Mills gauge theory related by a "large" gauge transformation are topologically inequivalent, since their so-called Chern-Simons numbers are different. The Chern-Simons number n_{CS} associated with potential $\mathbf{A}_\mu$ is defined as follows:

$$n_{CS}\left(\mathbf{A}_\mu\right) \equiv \frac{1}{16\pi^2}\int d^3\mathbf{x}\epsilon^{ijk}\left(A_i^a\partial_j A_k^a + \frac{1}{3}\epsilon^{abc}A_i^a A_j^b A_k^c\right) \tag{10}$$

and if $\mathbf{A}_\mu''$,$\mathbf{A}_\mu'$ are related by a "large" gauge transformation of the form (3) with winding number n, then $n_{CS}\left(\mathbf{A}_\mu''\right) = n_{CS}\left(\mathbf{A}_\mu'\right) + n$. But in a semi-classical treatment, quantum tunneling between them is possible through quantum tunneling. This suggests that the classical vacua are indeed distinct, and that a "large" gauge transformation represents a change from one physical situation to another. If so, symmetry under "large" gauge transformations is not just a theoretical symmetry but reflects an empirical symmetry of a non-Abelian Yang-Mills gauge theory. This favors the first analogy.

But Rubakov then goes on to offer an alternative (but allegedly equivalent!) perspective, when he says (on page 277)

> From the point of view of gauge-invariant quantities, topologically distinct classical vacua are equivalent, since they differ only by a gauge transformation. Let us identify these vacua. Then the situation becomes analogous to the quantum-mechanical model of the pendulum.

From this perspective, even "large" gauge transformations lead from a single classical vacuum state back into an alternative representation of that same state! Is this perspective legitimate? If it is, how can it be equivalent to a view according to which a "large" gauge transformation represents an empirical transformation between distinct states of a non-Abelian Yang-Mills gauge theory?

5 Are "Large" Gauge Transformations Empirical?

Consider first a purely classical non-Abelian Yang-Mills gauge theory. If it has models that represent distinct degenerate classical vacua, what is the physical difference between these vacua? Models related by a "large" gauge transformation are characterized by different Chern-Simons numbers, and one might take these to exhibit a difference in the intrinsic properties of situations they represent. But it is questionable whether the Chern-Simons number of a gauge configuration represents an intrinsic property of that configuration, even if a *difference* in Chern-Simons number represents an intrinsic *difference* between gauge configurations. Perhaps Chern-Simons numbers are like velocities in models of special relativity. The velocity assigned to an object in a model of special relativity does not represent an intrinsic property of that object, even though that theory does distinguish in its models between situations involving objects moving with different *relative* velocities. It was this latter distinction that proves critical to establishing that Lorentz boosts are empirical symmetries of situations in a special relativistic world.

So does a *difference* in Chern-Simons number represent an intrinsic *difference* between classical vacua in a purely classical non-Abelian Yang-Mills gauge theory? There is no reason to believe that it does. For it to do so, the theory would have to include models representing *more than one* vacuum state at once, where the distinct vacua are represented by different Chern-Simons numbers in *every* such model. Such distinct vacua extend over all space. So they could all be represented within a single model only if it represented them as occurring at different times. But topologically distinct vacua are separated by an energy barrier, and in the purely classical theory this cannot be overcome. So there is no representation within a single model of the purely classical theory of vacua with different Chern-Simons numbers. There is no reason to believe that a "large" gauge transformation represents an empirical transformation between distinct vacuum states of a purely classical non-Abelian Yang-Mills gauge theory.

According to a semi-classical theory, vacua with different Chern-Simons numbers *can* be connected by tunnelling through the potential barrier that separates them. So such a theory can model a single situation involving more than one such vacuum state, each obtaining at a different time. Moreover, no model of this theory represents these states as having the *same* Chern-Simons numbers. Perhaps this justifies the conclusion that in a world truly described by such a theory a "large" gauge transformation *would* represent an empirical transformation between distinct vacuum states. But we do not live in such a world.

The $\theta-$vacuum of a fully quantized non-Abelian Yang-Mills gauge theory is non-degenerate and symmetric under "large" as well as "small" gauge transformations. Analogies with the periodic potential and quantum pendulum suggest that it be expressed in the form

$$|\theta\rangle = \sum_{n=-\infty}^{+\infty} \exp\{-in\theta\} |n\rangle \tag{11}$$

where state $|n\rangle$ corresponds to a classical state with Chern-Simons number n. But not only the $\theta-$vacuum but the whole theory is symmetric under "large" gauge transformations. So a generator $\hat{U}$ of "large" gauge transformations commutes not only with the Hamiltonian but with all observables. It acts as a so-called "superselection operator" that separates the large Hilbert space of states into distinct superselection sectors, between which no superpositions are possible. Physical states are therefore restricted to those lying in a single superselection sector of the entire Hilbert space. Hence every physical state of the theory, including $|\theta\rangle$, is an eigenstate of $\hat{U}$.

Now there is an operator $\hat{U}_1$ corresponding to a "large" gauge transformation with winding number 1,

$$\hat{U}_1 |n\rangle = |n+1\rangle \tag{12}$$

from which it follows that none of the states $|n\rangle$ is a physical state of the theory! This theory cannot model situations involving *any* state corresponding to a classical vacuum with definite Chern-Simons number, still less a situation involving two or more states corresponding to classical vacua with *different* Chern-Simons numbers. Consequently, "large" gauge transformations in a fully quantized non-Abelian Yang-Mills gauge theory do *not* represent physical transformations, and symmetry under "large" gauge transformations is not an empirical symmetry. There is no difference in this respect between "large" and "small" gauge transformations.

6 Are They Really Gauge Transformations?

There are several reasons why it remains important to better understand the difference between "large" and "small" gauge transformations. One reason is that doing so will help to resolve the following apparent paradox.

Two beliefs are widely shared. The first belief is that "local" gauge transformations implement no empirical symmetry and therefore have no direct empirical consequences. The second belief is that "*global*" gauge transformations have *indirect* empirical consequences *via* Noether's Theorem, including the conservation of electric charge. The paradox arises when one notes that a "*global*" gauge transformation appears as a special case of a "local" gauge transformation. If "local" gauge symmetry is a purely formal symmetry, how can (just) this special case of it have even *indirect* empirical consequences?

Another reason is to appreciate why some (e.g. Domenico Giulini) have proposed that we make

> a clear and unambiguous distinction between proper physical symmetries on one hand, and gauge symmetries or mere automorphisms of the mathematical scheme on the other (Giulini(2003), p.289)

The proposed distinction would classify invariance under "small" gauge symmetries as a gauge symmetry, but invariance under "large" gauge transformations as a proper physical symmetry. It is founded on an analysis of gauge in the framework of constrained Hamiltonian systems.

The guiding principle is to follow Dirac's proposal by identifying gauge symmetries as just those transformations on the classical phase-space representation of the state of such a system that are generated by its first-class constraint functions. In a classical Yang-Mills gauge theory, these are precisely those generated by the so-called Gauss constraint functions, such as the function on the left-hand side of equation

$$\boldsymbol{\nabla}.\mathbf{E} = 0 \tag{13}$$

in the case of pure electromagnetism.

Giulini(2003) applies this principle to a quantized Hamiltonian system representing an isolated charge distribution in an electromagnetic field. He concludes that the gauge symmetries of this system consist of all and only those "local" gauge transformations on the quantized fields that leave unchanged both the asymptotic electromagnetic gauge potential $\hat{A}_\mu$ and the distant charged matter field. A "global" gauge transformation corresponding to a constant phase rotation in the matter field does *not* count as a gauge symmetry since it is not generated by the Gauss constraint (or any other first-class constraint) function. Rather, "global" $U(1)$ phase transformations would be associated with what Giulini calls *physical* symmetries. According to Giulini(2003) (p.308)

> This is the basic and crucial difference between local and global gauge transformations.

The formalism represents the charge of the system dynamically by an operator $\hat{Q}$ that generates translations in a coordinate corresponding to an additional degree of freedom on the boundary in the dynamical description. A charge superselection rule, stating that all observables commute with the charge operator, is equivalent to the impossibility of localizing the system in this new coordinate. Consequently, conservation of charge implies that translations in this additional degree of freedom count as physical symmetries for Giulini. So conservation of charge is equivalent both to the existence of these symmetries, and (by Noether's first theorem) to the "global" gauge symmetry of the Lagrangian. But these physical symmetries do not correspond to gauge symmetries, either "global" or "local", since they affect neither the gauge potential nor the phase of the matter field.

It is hard to argue that these novel physical symmetries are empirical. No operational procedures are specified to permit measurement of the additional degrees of freedom, and these attach on a boundary which is eventually removed arbitrarily far away. But even if such a new physical symmetry were empirical, it would not correspond to any constant phase change. A "global" gauge symmetry would still not entail any corresponding empirical symmetry.

This delicate relation between "global" gauge transformations and some other physical symmetry helps to resolve the apparent paradox outlined above. A"global" gauge transformation is not merely a special case of a "local" gauge transformation. Indeed, the constrained Hamiltonian approach provides a valuable perspective from which it is not even appropriately classified as a gauge transformation.

This perspective illuminates the distinction between "large" and "small" gauge transformations more generally. As Giulini put it in 1995, in Yang-Mills theories

> it is the Gauss constraint that declares some of the formally present degrees of freedom to be physically nonexistent. But it only generates the identity component of asymptotically trivial transformations, leaving out the long ranging ones which preserve the asymptotic structure imposed by boundary conditions as well as those not in the identity component of the asymptotically trivial ones. These should be considered as proper physical symmetries which act on physically existing degrees of freedom.

Whether the constrained Hamiltonian approach to gauge symmetry establishes that "large" gauge transformations correspond to empirical symmetries is more sensitive to theoretical context than Giulini's last sentence seems to allow. But the approach certainly shows that not only a "global" gauge transformation but any "large" gauge transformation not generated by a Gauss constraint is very different from the "local" gauge symmetries that it does generate.

7 The $\theta-$Vacuum in a Loop Representation

The availability of loop representations of quantized Yang-Mills theories has interesting implications for the nature of the $\theta-$vacuum. Recall that when the theory is non-Abelian, "large" gauge transformations with non-zero winding number connect potential states with different Chern–Simons numbers, including different candidates for the lowest-energy, or vacuum, state of the field. Requiring that the theory be symmetric under such "large" gauge transformations implies that the actual vacuum state is a superposition of all these candidate states of the form

$$|\theta\rangle = \sum_{n=-\infty}^{+\infty} \exp\{-in\theta\} \, |n\rangle \tag{14}$$

where θ is an otherwise undetermined parameter—a fundamental constant of nature.

Associated with the $\theta-$vacuum is an additional term proportional to $\epsilon_{\mu\nu\rho\sigma}F^{a\mu\nu}F^{a\rho\sigma}$ that enters the effective Lagrangian density for quantum chromodynamics

$$\mathcal{L}_{QCD} = \overline{\psi}_a(i\gamma^\mu D_\mu - m)\psi^a - \frac{1}{4}F_{a\mu\nu}F^{a\mu\nu} + \frac{\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}F^{a\mu\nu}F^{a\rho\sigma} \tag{15}$$

—unless the value of θ is zero, in which case this term itself becomes zero. It turns out that certain empirical consequences of quantum chromodynamics are sensitive to the presence of this extra term: if it were present, then strong interactions would violate two distinct discrete symmetries, namely parity and charge conjugation symmetry. Experimental tests have shown that $|\theta| \leq 10^{-10}$, making one suspect that in fact $\theta = 0$. This fact—that of all the possible real number values it could take on, θ appears to be zero—is known as the *strong CP problem*. Various solutions have been offered, several of which appeal to some new physical mechanism that intervenes to force θ to equal 0. But from the perspective of a loop representation, there is no need to introduce θ as a parameter in the first place. I quote (Fort and Gambini, 2000):

> It is interesting to speculate what would happen if from the beginning holonomies were used to describe the physical interactions instead of vector potentials. Probably we would not be discussing the strong CP problem. This would simply be considered as an artifact of an overdescription of nature, by means of gauge potentials, which is still necessary in order to compute quantities by using the powerful perturbative techniques. From this perspective, the strong CP problem is just a matter of how we describe nature rather than being a feature of nature itself. (p.348)

As Fort and Gambini explain, when a theory is formulated in a loop/path representation, all states and variables are automatically invariant under both "small" and "large" gauge transformations, so there is no possibility of introducing a parameter θ (as in equation (11)) to describe a hypothetical superposition of states that are not so invariant. While the conventional perspective makes one wonder why θ should equal zero, from the loop perspective there is no need to introduce any such parameter in the first place. Once formulated, the loop representation will be equivalent to the usual connection representation with $\theta = 0$.

One can introduce an arbitrary parameter θ into a loop representation of a more complex theory, as Fort and Gambini show. But from the holonomy perspective there would have been no empirical reason to formulate such a more complex theory, and the fact that even more precise experiments do not require it would be a considered a conclusive reason to prefer the simpler theory—the one that never introduced an empirically superfluous θ parameter.