Gauge Symmetries in the Standard Model

Richard Healey,

University of Arizona

rhealey@email.Arizona.edu

Vancouver, May 23rd 2007

1 Gauge Symmetry in Quantum Theories

2 The Gauge Argument

3 Non-Abelian Gauge Theories

4 The $\theta-$Vacuum

5 Two Analogies

6 Are "Large" Gauge Transformations Empirical?

1 Gauge Symmetry in Quantum Theories

A gauge transformation in classical electromagnetism involves only the electric scalar and magnetic vector potentials ϕ, A; or their Lorentz-covariant combination A_μ, the 4-vector electromagnetic potential. It takes the form

$$A_\mu(x) \to A_\mu(x) + \partial_\mu \Lambda(x) \tag{1}$$

for arbitrary suitably differentiable function $\Lambda(x)$ (here x is an abbreviated label for the point with coordinates (t, x, y, z)).

The state of matter may be represented in quantum mechanics by a wave-function or quantum field $\psi(x)$. In a quantum particle theory the values of ψ are complex numbers, or $n-$tuples of complex numbers. In a quantum field theory the values of ψ are linear operators on a vector space (a Hilbert space), or $n-$tuples of them: a vector from this space represents the state of the field.

In a quantum theory, an electromagnetic gauge transformation transforms not only A_μ but also ψ: These transform are as follows

$$
\begin{aligned}
A_\mu(x) &\rightarrow A_\mu(x) + \partial_\mu \Lambda(x) \quad \text{(2a)} \ \textbf{Gauge} \\
\psi(x) &\rightarrow \exp[-(ie)\Lambda]\psi(x) \quad \text{(2b)} \ \textbf{Transformation}
\end{aligned}
$$

where $i = \sqrt{-1}$, and e is the charge of the particles or field quanta. (With natural units in which $\hbar = c = 1$.)

If Λ is a constant, A_μ is unaffected and ψ is modified by a constant overall phase factor: This is called a *global* gauge transformation.

If $\Lambda(x)$ varies with space-time position x, both ψ and A_μ are modified: This is called a *local* gauge transformation.

A quantum theory possesses electromagnetic *gauge symmetry* if and only if all gauge transformations are theoretical symmetries of its models.

A quantum particle theory is trivially symmetric under global gauge transformations, since wave-functions that differ only by an overall phase-factor represent the same quantum state.

Global gauge symmetry of a quantum field theory is *not* trivial: here global electromagnetic gauge symmetry is equivalent (by Noether's first theorem) to conservation of electric charge.

The so-called *gauge argument* attempts to derive the existence and properties of interactions involving matter by extending the global gauge symmetry of a quantum field theory of matter to a symmetry under *local* gauge transformations. This would be one way to derive the existence and properties of electromagnetic interactions—if these were not already known! But electromagnetic gauge symmetry may be generalized to other kinds of gauge symmetry. Here, application of the gauge argument might not only explain but also predict the existence and properties of *new* interactions.

3 Non-Abelian Gauge Theories

Even though it is invalid, the gauge argument does provide a unified heuristic perspective on the interactions of the Standard Model of elementary particles. Electromagnetism was only the first interaction to be described by a locally gauge-symmetric theory in that Model. Indeed, the electromagnetic interaction is now regarded as only one aspect of a unified electro-weak interaction that also includes weak interactions. A third gauge theory, quantum chromodynamics, models the strong interaction.

Each of these new theories is said to be *non-Abelian*, since the generalized gauge transformations it prescribes are represented by elements of groups some of whose elements g_1, g_2 fail to commute ($g_1 \cdot g_2 \neq g_2 \cdot g_1$). The group $U(1)$, by contrast, is an Abelian group. The group $SU(2)$ was the first non-Abelian group whose generalized gauge transformations were investigated, by Yang and Mills in 1954. All the gauge theories of the Standard Model are now known as Yang-Mills theories.

In an $SU(2)$ Yang-Mills gauge theory, the generalization of the electromagnetic potential is represented by a Lorentz-covariant magnitude $\mathbf{A}_\mu$, each of whose components is itself a 2-dimensional square matrix with complex number coefficients. In one representation this is given by

$$\mathbf{A}_\mu = A_\mu^j \frac{\sigma_j}{2i} \tag{6}$$

where σ_j $(j = 1, 2, 3)$ are Pauli spin matrices and $i = \sqrt{-1}$. Corresponding to this generalized potential is a generalized field strength $\mathbf{F}_{\mu\nu}$, where

$$\mathbf{F}_{\mu\nu} = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu + [\mathbf{A}_\mu, \mathbf{A}_\nu] \tag{7}$$

and $[\mathbf{A}_\mu, \mathbf{A}_\nu] \equiv \mathbf{A}_\mu \mathbf{A}_\nu - \mathbf{A}_\nu \mathbf{A}_\mu$. The third, non-linear term in a non-Abelian theory marks an important difference between (7) and the analogous equation (5) from electromagnetism.

The Abelian equation (2a) generalizes to

$$\mathbf{A}_\mu \to \mathbf{A}'_\mu = \mathbf{U}\mathbf{A}_\mu \mathbf{U}^\dagger + (\partial_\mu \mathbf{U})\mathbf{U}^\dagger \tag{8}$$

where $\mathbf{U}$ is an element of $SU(2)$, and $\mathbf{U}^\dagger$ is the adjoint matrix. A further interesting difference from the Abelian case is that the non-Abelian field strength $\mathbf{F}_{\mu\nu}$ is not gauge-invariant but gauge covariant:

$$\mathbf{F}_{\mu\nu} \to \mathbf{U}\mathbf{F}_{\mu\nu}\mathbf{U}^\dagger \tag{9}$$

4 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 supposedly connect models of distinct but indistinguishable situations. If this is so, it shows 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

$$S = \frac{1}{2g^2}\int Tr(\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu})d^4x \qquad (10)$$

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

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

$$\mathbf{A}_\mu \rightarrow \mathbf{A}'_\mu = \mathbf{U}\mathbf{A}_\mu\mathbf{U}^\dagger + (\partial_\mu \mathbf{U})\mathbf{U}^\dagger,\ \mathbf{F}_{\mu\nu} \rightarrow \mathbf{U}\mathbf{F}_{\mu\nu}\mathbf{U}^\dagger \qquad (11)$$

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.

$$\mathbf{A}_\mu = 0 \rightarrow \mathbf{A}'_j(\mathbf{x}) = \{\partial_j \mathbf{U}(\mathbf{x})\}\mathbf{U}^\dagger(\mathbf{x}),\ \mathbf{A}'_0 = 0 \quad (12)$$

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

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.

5 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

$$|\theta\rangle = \sum_{n=-\infty}^{+\infty} \exp\{-in\theta\} |n\rangle \tag{13}$$

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

$$\hat{H} = \frac{1}{2m}[-i(\boldsymbol{\nabla} - ie\mathbf{A})]^2 + V \tag{14}$$

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

$$\hat{H} = -\frac{1}{2ml^2}\left(\frac{d}{d\omega} - ielA\right)^2 + V(\omega) \tag{15}$$

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

$$\psi(\omega) = \exp\left[ie\int_0^\omega lAd\omega'\right]\varphi(\omega) \tag{16}$$

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

$$\hat{H}_\varphi = -\frac{1}{2m}\frac{d^2}{d\omega^2} + V(\omega) \tag{17}$$

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

$$\varphi(\omega + 2\pi) = \exp\{-ie\Phi\}\varphi(\omega) \qquad (18)$$

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:

$$n_{CS}\left(\mathbf{A}_\mu\right) \equiv \frac{1}{16\pi^2}\int d^3\mathbf{x}\epsilon^{ijk}\left(A_i^a\partial_j A_k^a + \frac{1}{3}\epsilon^{abc}A_i^a A_j^b A_k^c\right) \tag{19}$$

and if $\mathbf{A}''_\mu$,$\mathbf{A}'_\mu$ are related by a "large" gauge transformation of the form (12) 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?

6 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 is 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

$$|\theta\rangle = \sum_{n=-\infty}^{+\infty} \mathsf{exp}\{-in\theta\} \, |n\rangle \tag{20}$$

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,

$$\hat{U}_1 \, |n\rangle = |n+1\rangle \tag{21}$$

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.

2 The Gauge Argument

Suppose one were to try to develop a quantum field theory for a charged matter field. One might proceed to quantize a classical field theory by promoting classical field magnitudes into operators obeying suitable generalizations of the Heisenberg commutation relations. Given Noether's first theorem, the empirical fact of charge conservation is naturally incorporated into such a classical field theory by taking the field to be represented by a complex-valued function $\psi(x)$ that transforms as

$$\psi(x) \rightarrow \exp[-(ie)\Lambda]\psi(x) \qquad (3)$$

under a global gauge transformation. This transformation is generated by an element of the group $U(1)$, and acts as a rotation of the "arrow" representing the value of ψ in the complex plane—indeed, the *same* rotation at every space-time point x.

But now one might be struck by the thought that no physical content has yet been given to the operation of comparing the "directions" in which that arrow is pointing at different space-time points—this is not a direction in *physical* space(-time). What is required is a *connection* that specifies how to compare directions in this "internal" group space. The actual values $\psi(x)$ are therefore unimportant: what matters is how these are related to one another *via* this connection. To allow for this arbitrariness, (3) should be generalized to

$$\psi(x) \rightarrow \exp[-(ie)\Lambda(x)]\psi(x) \tag{4}$$

and the theory should be required to be locally as well as globally gauge symmetric.

The Dirac equation governing a classical electron field is *not* locally gauge symmetric. But it may be *made* locally gauge symmetric by replacing the ordinary derivative operator ∂_μ by the so-called *covariant derivative* operator $\partial_\mu - ieA_\mu$ where $A_\mu(x)$ is a newly introduced Lorentz-covariant field representing the desired connection. This satisfies (2a), and so it is tempting to identify it with the electromagnetic 4-vector potential.

If this $A_\mu(x)$ is indeed a new field, then it has its own dynamics. The simplest dynamical equations derivable from an action principle applied to a Lorentz and locally gauge-invariant Lagrangian for the new field have the form of Maxwell's equations for a source-free electromagnetic field represented by the Lorentz tensor $F_{\mu\nu}$, where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{5}$$

Moreover, application of an action principle to the total Lagrangian yields Maxwell's equations with the usual charged matter-field current as source. So we seem to have derived the existence and properties of electromagnetism, including its interactions with matter, just by extending the global gauge symmetry of the matter field to local gauge symmetry!

But this argument is invalid. Its plausibility rests on two implicit assumptions which need to be added explicitly to render it sound.

The first assumption is that the connection represented by $A_\mu(x)$ need not be *flat*—that the curvature given by (5) may not be zero. The second assumption is that this curvature is not fixed but dynamic, and indeed depends in a specific way on the charged matter current. It is this second assumption that justified the application of action principles to derive the dynamics of the field $F_{\mu\nu}$, in the presence as well as in the absence of the matter current. While these assumptions turn out to be true, they certainly don't follow from the requirement of local gauge symmetry. Local gauge symmetry remains a purely formal symmetry—a theoretical symmetry of the matter field which, by itself, implies no corresponding empirical symmetry. Its role in the gauge argument is purely heuristic—to suggest a possibility that nature turns out to have realized.

7 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

$$\nabla . \mathbf{E} = 0 \tag{22}$$

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 *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: these correspond to physical symmetries. 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. But a translation in this additional coordinate is quite distinct from a transformation of the form

$$\begin{aligned} A_\mu &\rightarrow A_\mu + \partial_\mu \Lambda \\ \psi &\rightarrow \exp[-(ie)\Lambda]\psi \end{aligned} \tag{23}$$

Now if this charge superselection rule is assumed to be strict, then "global" $U(1)$ phase transformations are forbidden, and there is no associated physical symmetry. So conservation of charge implies that no physical symmetry is associated with a "global" gauge transformation at the same time that it is equivalent (by Noether's Theorem) to "global" gauge symmetry of the Lagrangian! It is only in a theory that permitted states of different charge to be superposed that "global" gauge transformations would be manifested as empirical symmetries; and even in such a theory these would not be symmetries of the form of equation (23).

This delicate relation between "global" gauge transformations and some associated empirical 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 seems more sensitive to theoretical context than Giulini's last sentence would allow. But it 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.

8 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

$$|\theta\rangle = \sum_{n=-\infty}^{+\infty} \mathrm{exp}\{-in\theta\}\,|n\rangle \tag{24}$$

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

$$\begin{aligned}\mathcal{L}_{QCD} &= \overline{\psi}_a(i\gamma^\mu D_\mu - m)\psi^a - \frac{1}{4}F_{a\mu\nu}F^{a\mu\nu} \\ &\quad + \frac{\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}F^{a\mu\nu}F^{a\rho\sigma}\end{aligned} \tag{25}$$

—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 (20)) 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.