Symmetries, Symmetry Breaking, Gauge Symmetries*

F. Strocchi
INFN, Sezione di Pisa, Pisa, Italy

Abstract

The concepts of symmetry, symmetry breaking and gauge symmetries are discussed, their operational meaning being displayed by the observables and the (physical) states. For infinitely extended systems the states fall into physically disjoint phases characterized by their behavior at infinity or boundary conditions, encoded in the ground state, which provide the cause of symmetry breaking without contradicting Curie Principle. Global gauge symmetries, not seen by the observables, are nevertheless displayed by detectable properties of the states (superselected quantum numbers and parastatistics). Local gauge symmetries are not seen also by the physical states; they appear only in non-positive representations of field algebras. Their role at the Lagrangian level is merely to ensure the validity on the physical states of local Gauss laws, obeyed by the currents which generate the corresponding global gauge symmetries; they are responsible for most distinctive physical properties of gauge quantum field theories. The topological invariants of a local gauge group define superselected quantum numbers, which account for the $\theta$ vacua.

*Talk at the Triennial International Conference "New Developments in Logic and Philosophy of Science", Rome 18-20 June, 2014
1 Introduction

The concepts of symmetries, symmetry breaking and gauge symmetries, at the basis of recent developments in theoretical physics, have given rise to discussions from a philosophical point of view. Critical issues are the meaning of spontaneous symmetry breaking (appearing in conflict with the Principle of Sufficient Reason) and the physical or operational meaning of gauge symmetries.

The aim of this talk is to offer a revisitation of the problems strictly in terms of operational considerations. The starting point (not always emphasized in the literature) is the realization that the description of a physical system involves both the observables, identified by the experimental apparatuses used for their measurements, and the states, which define the experimental expectations. Since the protocols of preparations of the states may not always be compatible, i.e. obtainable one from the other by physically realizable operations, the states fall into disjoint families, called phases, corresponding to incompatible realizations of the system. This is typically the case for infinitely extended systems, where different behavior or boundary conditions of the states at space infinity identify disjoint phases due to the inevitable localization of any realizable operation.

This feature, which generically is not shared by finite dimensional systems, provides the explanation of the phenomenon of spontaneous symmetry breaking, since the boundary conditions at infinity encoded in the ground state represent the cause of the phenomenon in agreement with Curie principle.

The role of the states is also crucial for the physical meaning of gauge symmetries, which have been argued to be non-empirical because they are not seen by the observables. The fact that non-empirical constituents may characterize the theoretical description of subnuclear systems, as displayed by the extraordinary success of the standard model of elementary particle physics, has provoked philosophical discussion on their relevance (see [1]). For the discussion of this issue it is important to distinguish global (GGS) and local gauge symmetries (LGS).

The empirical consequences of the first is displayed by the properties of the states, since invariant polynomials of the gauge generators define

\footnote{An updated and comprehensive account may be found in [1].}
elements of the center of the algebra of observables $\mathcal{A}$, whose joint spectrum labels the representations of $\mathcal{A}$ defining superselected quantum numbers; another empirical consequence of a global gauge group is the parastatistics obeyed by the states. Actually the existence of a gauge group can be inferred from such properties of the states.

At the quantum level, the group of local gauge transformations connected to the identity may be represented non-trivially only in unphysical non-positive representations of the field algebra and therefore they reduce to the identity not only on the observables, but also on the physical states.

From a technical point of view, a role of LGS is to identify (through the pointwise invariance under them) the local observable subalgebras of auxiliary field algebras (represented in non-positive representations). LGS also provide a useful recipe for writing down Lagrangians which automatically lead to the validity on the physical states of local Gauss laws (LGL), satisfied by the currents which generate the corresponding GGS. Actually, LGL appear as the important physical counterpart of LGS representing the crucial distinctive features of Gauge QFT with respect to ordinary QFT.

A physical residue of LGS is also provided by their local topological invariants, which define elements of the center of the local algebras of observables, the spectrum of which label the inequivalent representations corresponding to the so-called $\theta$ vacua. The occurrence of such local topological invariants explains in particular the breaking of chiral symmetry in Quantum Chromodynamics (QCD), with no corresponding Goldstone bosons.

Finally, since only observables and states (identified by their expectations of the observables [2] [3]) are needed for a complete description of a physical system, and both have a deterministic evolution, the claimed violation of determinism in gauge theories is completely unjustified from a physical and philosophical point of view.
2 Symmetries and symmetry breaking

For the clarification of the meaning and consequences of symmetries in physics, from the point of view of general philosophy, a few basic concepts are helpful.

Quite generally, the description of a physical system (not necessarily quantum!) is (operationally) given \[2\] \[3\] in terms of

1) the observables, i.e. the set of measurable quantities of the system, which characterize the system (and generate the so-called algebra $\mathcal{A}$ of observables)

2) their time evolution

3) the set $\Sigma$ of physical states $\omega$ of the system, operationally defined by protocols of preparations and characterized by their expectations of the observables $\{\omega(A), A \in \mathcal{A}\}$

Operationally, an observable $A$ is identified by the actual experimental apparatus which is used for its measurement, (two apparatuses being identified if they yield the same expectations on all the states of the system)

The first relevant point is the compatible realization of two different states, meaning that they are obtainable one from the other by physically realizable operations. This defines a partition of the states into physically disjoint sets, briefly called phases, with the physical meaning of describing disjoint realizations of the system, like disjoint thermodynamical phases, disjoint worlds or universes.

For infinitely extended systems, in addition to the condition of finite energy, a very strong physical constraint is that the physically realizable operations have inevitably some kind of localization, no action at space infinity being physically possible. Thus, for the characterization of the states of a phase $\Gamma$, a crucial role is played by their large distance behavior or by the boundary conditions at space infinity, since they cannot be changed by physically realizable operations. Typically, such a behavior at infinity of the states of a given phase $\Gamma$ is codified by the lowest energy state or ground state $\omega_0 \in \Gamma$, all other states of $\Gamma$ being describable as “localized” modifications of it. Thus, $\omega_0$ identifies $\Gamma$ and defines a corresponding (GNS) representation $\pi_\Gamma(\mathcal{A})$ of the observables in a Hilbert space $\mathcal{H}_\Gamma$, with the cyclic ground state vector $\Psi_0^\Gamma$.

\[2\]This point is discussed for both classical and quantum systems in [4], [5].
The simplest realization of symmetries is as transformations of the observables commuting with time evolution, operationally corresponding to the transformations of the experimental apparatuses which identify the observables (e.g., translations, rotations). This is more general than Wigner definition of symmetries as transformations of the states which leave the transition probabilities invariant (adapted to the case of the unique Schrödinger phase of atomic systems).

Actually, the disentanglement of symmetry transformations of the observables (briefly algebraic symmetries) from those of the states (Wigner symmetries), is the crucial revolutionary step at the basis of the concept of spontaneous symmetry breaking, which comes into play when there is more than one phase.

An algebraic symmetry $\beta$ defines also a symmetry of the states of a phase $\Gamma$ (i.e., a Wigner or unbroken symmetry) iff it may be represented by unitary operators $U_\beta$ in $\mathcal{H}_\Gamma$.

An algebraic symmetry $\beta$ always defines a symmetry of the whole set of states $\Sigma$:

$$\omega \rightarrow \beta^* \omega \equiv \omega_\beta, \quad \omega_\beta(A) \equiv \omega(\beta^{-1}(A)), \quad \forall A \in \mathcal{A}, \quad (2.1)$$

but in general $\omega$ and $\omega_\beta$ need not to belong to the same phase $\Gamma$, i.e. their preparation may not be compatible, so that the symmetry $\beta$ cannot be experimentally displayed in $\Gamma$ as invariance of transition probabilities, by means of physically compatible operations (spontaneously broken symmetry). Thus, the breaking of $\beta$ in $\Gamma$ is characterized by the existence of states $\omega \in \Gamma$ (typically the ground or vacuum state $\omega_0$) such that $\omega_\beta \notin \Gamma$.

The philosophical issue of symmetry breaking, also in connection with Curie principle, has been extensively debated often with misleading or wrong conclusions.

A widespread opinion is that symmetry breaking occurs whenever the ground state is not symmetric, but this is not correct for finite systems, for which (under general conditions) there is only one (pure) phase $\Gamma$, so that both $\omega_0$ and $\omega_{0,\beta}$ belong to $\Gamma$ and $\beta$ is described by a unitary operator.

Thus, the finite dimensional (mechanical) models, widely used in the literature to illustrate spontaneous symmetry breaking, on the basis of the existence of non-symmetric ground states, are conceptually
On the other hand, for a pure phase of an infinitely extended system, thanks to the uniqueness of the translationally invariant state (implied by the cluster property which characterizes pure phases), the non-invariance of the ground state $\omega_0 \in \Gamma$ under an internal symmetry $\beta$ (i.e. commuting with space-time translations) implies that $\omega_{0,\beta}$ cannot belong to $\Gamma$ and $\beta$ is broken in $\Gamma$. Under these conditions, the non-invariance of the ground state provides an explanation in agreement with Curie principle, identifying the cause in non-symmetric boundary conditions at infinity encoded in the ground state (see [4] pp.23, 102). The philosophically deep loss of symmetry requires the existence of disjoint realizations of the system, which is related to its infinite extension.

The existence of an algebraic symmetry reflects on empirical properties of the states and may be inferred from them. In fact, an unbroken symmetry implies the validity of Ward identities, which codify the existence of conserved quantities and of selection rules satisfied by the states; for continuous symmetries the conservation laws hold even locally by the existence of current continuity equations implied by the first Noether theorem ([5], p.146-7). For a continuous symmetry group $G$ broken in $\Gamma$, even if the generators do not exist as operators in $\mathcal{H}_\Gamma$, the existence of a representation of $G$ at the algebraic level, ([4], Chapter 15), implies symmetry breaking Ward identities which display corrections given by non-symmetric ground state expectations, called non-symmetric order parameters; an important empirical consequence is the existence of Goldstone bosons, for sufficiently "local" dynamics ([4], Chapters 15-17).

---

3 The standard models are a particle in a double well or in a mexican hat potential (see also [6], [7]). The example of an elastic bar on top of which a compression force is applied, directed along its axis, exhibits a continuous family of symmetry breaking ground states, but spontaneous symmetry breaking occurs only in the limit of infinite extension of the bar; otherwise, both in the classical as well in the quantum case, there is no obstruction for reaching one ground state from any other.
3 Global gauge symmetries

For the debated issue of the empirical meaning of global gauge symmetries (GGS) (which by definition act trivially on the observables), a crucial (overlooked) point is that a complete description of a physical system involves both its algebra of observables and the states or representations which describe its possible phases. In fact, even if there is no (non-trivial) transformation of the observables corresponding to GGS, GGS are strictly related to the existence of disjoint representations of the observable algebra and their empirical meaning is to provide a classification of them in terms of superselected quantum numbers [8]. This is clearly illustrated by the following examples.

Example 1. Consider a free massive fermion field $\psi$ transforming as the fundamental representation of an internal $U(2) = U(1) \otimes SU(2)$ symmetry with the algebra of observables defined by its pointwise invariance under $U(2)$. The existence of the (free) Hamiltonian selects the Fock representation in $H_F$ for the field algebra $F$ generated by $\psi$ and this implies the existence of the generator $N$ of $U(1)$ and of the Casimir invariant

$$T^2 \equiv \sum_{\alpha=1}^{3}(Q^\alpha)^2, \quad Q^\alpha \equiv \int d^3 x \psi^* (x) T^\alpha \psi (x), \quad (3.1)$$

with $T^\alpha$, $\alpha = 1, \ldots, 3$, the representatives of the generators of $SU(2)$. $N$ and $T^2$ are invariant under the gauge group $U(2)$ and as such they (or better their exponentials $U_N (\alpha) = \exp i\alpha N$, $U_T (\beta) = \exp i\beta T^2$, $\alpha, \beta \in \mathbb{R}$) may be taken as elements of the center $Z$ of the observable algebra $A$. The eigenvalues $n \in \mathbb{N}$ of $N$ and $j(j+1)$ ($j \in \frac{1}{2} \mathbb{N}$) of $T^2$ label the representations of $A$ in $H_F$ and the fermion fields $\psi^*$, $\psi$ act as intertwiners between the inequivalent representations of $A$, by increasing/decreasing the numbers $n$ and $j$.

Had we started by considering only the observable algebra $A$, we would have found that its representations are labeled by the (superselected) quantum numbers $n$ and $j(j+1)$, corresponding to the spectrum of the central elements $U_N (\alpha)$, $U_T (\beta)$ and that the state vectors of the representations of $A$ are obtained by applying intertwiners to the $n = 0$, $j = 0$ representation, consisting of the Fock vacuum.
We would then be led to consider a larger (gauge dependent) algebra $\mathcal{F}$ generated by the intertwiners, to interpret $n$ as the spectrum of the generator $N$ of a $U(1)$ group and to infer the existence of an $SU(2)$ group with $j(j+1)$ the eigenvalues of the associated $T^2$. Such a reconstructed $U(2)$ group acts non-trivially on the intertwiners, but trivially on the observables, namely is a global gauge group.

**Example 2.** A familiar physical system displaying the above structure is the quantum system of $N$ identical particles, even if in textbook presentations the relation between the gauge structure and the center of the observables is not emphasized.

The standard treatment introduces the (Weyl algebra $\mathcal{A}_W$ generated by the) canonical variables of $N$ particles and, by the very definition of indistinguishability, the observable algebra $\mathcal{A}$ is characterized by its pointwise invariance under the non-abelian group $\mathcal{P}$ of permutations, which is therefore a global gauge group.

As before, its role is that of providing a classification of the inequivalent representations of the observable algebra contained in the unique regular irreducible representation of $\mathcal{A}_W$, (equivalent to standard Schroedinger representation) in the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^N) \otimes \mathbb{C}$, where $\mathcal{P}$ is unbroken. $\mathcal{H}$ decomposes into irreducible representation of the observable algebra, each being characterized by a Young tableaux, equivalently by the eigenvalues of the characters $\chi_i$, $i = 1, \ldots, m$. For our purposes, the relevant point is that the characters are invariant functions of the permutations and, as such, may be considered as elements of the observable algebra, actually elements of its center $\mathcal{Z}$.

Thus, as before, the gauge group $\mathcal{P}$ provides elements of the center of the observables whose joint spectra label the representations of $\mathcal{A}$ defining superselected quantum numbers. Beyond the familiar one-dimensional representations (corresponding to bosons and fermions) there are higher dimensional representations, describing *parastatistics* (i.e. parabosons and parafermions).

Another empirical consequence of a global gauge group is the (observable) statistics obeyed by the states, a parastatistics of order $d$ arising as the result of an unbroken (compact) global gauge group acting on ordinary (auxiliary) bosons/fermions fields [10], [11]. In the model of Example 1, an observable consequence of the global gauge group $U(2)$ is that the corresponding particle states are parafermions.
of order two (meaning that not more than two particles may be in a state). The quarks have the properties of parafermions of order three as a consequence of the color group \( SU(3) \) (historically this was one of its motivations).

In conclusion, contrary to the widespread opinion that the gauge symmetries are not empirical, the global gauge symmetries are displayed by the properties of the states (superselected quantum numbers and parastatistics) and actually can be inferred from them.\(^4\)

It must be stressed that a global gauge symmetry emerges as an empirical property of a system by looking at the whole set of its different realizations; in a single factorial representation, the center of the observables is represented by a multiple of the identity and its physical meaning in terms of superselected quantum numbers is somewhat frozen. To reconstruct an operator of the center of \( \mathcal{A} \) one must look to its complete spectrum, i.e. to all factorial representations of \( \mathcal{A} \).

A continuous global gauge group becomes particularly hidden in those representations in which the exponentials of localized invariant polynomials of the generators converge to zero when the radius of the localization region goes to infinity. This corresponds to the case in which, in the conventional jargon, the global gauge group is broken.\(^5\)

In a representation \( \mathcal{H}_\Gamma \) of the field algebra in which the (continuous) gauge group \( G \) is broken, briefly called a \( G \)-broken representation, in contrast with the above examples, the charged fields do no longer intertwine between different representations of the observable algebra; in fact, they are obtainable as weak limits of gauge invariant fields in the Hilbert space \( \mathcal{H}_\Gamma \) (charge bleaching)\(^\text{12}\).

**Example 4.** The Bose-Einstein condensation is characterized by the breaking of a global \( U(1) \) gauge group (acting on the Bose particle field as the \( U(1) \) group of Example 1), as very clearly displayed by the free Bose gas.\(^6\) The \( U(1) \) breaking leads to the existence of Goldstone modes, the so-called Landau phonons, and the existence of such excitations may in turn indicate the presence of a broken \( U(1) \) symmetry. Finally, the gauge group is also reflected in the counting of the states.\(^6\)

\(^4\)The empirical meaning of the invariant functions of the generators of a global gauge group has been pointed out in \(^\text{5}\), pp.153-8 and later resumed by Kosso and others; (see also \(^\text{13}\), Chapter 7).

\(^5\)For a simple account see \(^\text{4}\), p. 106.
In $G$-unbroken representations of $\mathcal{A}$, to each irreducible representation of $G$ contained in the field algebra $\mathcal{F}$, there corresponds a single physical state, whereas in the fully broken case to each $d$-dimensional irreducible representation in $\mathcal{F}$, there correspond $d$ different physical states [14] (for a handy account see [5], Part B, Section 2.6).

4 Local gauge symmetries

Traditionally, a local gauge symmetry group is introduced as an extension of the corresponding global group $G$ by allowing the group parameters to become $C^\infty$ functions of spacetime. It is however better to keep distinct the local gauge group $\mathcal{G}$ parametrized by strictly localized functions (technically of compact support) from the corresponding global one $G$, since the topology of the corresponding Lie algebras is very different and invariance under $G$ does not imply invariance under $G$ (as displayed by the Dirac-Symanzik electron field, [13], p.159).

Also from a physical point of view, the two groups are very different, since in any (positive) realization (of the system) the group of local gauge transformations connected with the identity is represented trivially, whereas the global gauge group displays its physical meaning through the properties of the states (see the above examples). For example, the $U(1)$ global gauge group is non-trivially represented in Quantum Electrodynamics (QED) by the existence of the charged states, whereas the local $U(1)$ group reduces to the identity on the physical states ([13], Section 3.2).

Therefore, the natural question is which is the empirical meaning, if any, of a local gauge symmetry (LGS) $\mathcal{G}$ in QFT. From a technical point of view, pointwise invariance under $\mathcal{G}$ may be used for selecting the local subalgebra of observables, from an auxiliary field algebra $\mathcal{F}$, locality (strictly related to causality [14]) not being implied by $G$ invariance (e.g., in QED $\bar{\psi}(x)\psi(y)$ is invariant under $G = U(1)$, but not under $\mathcal{G}$ and is not a local observable field).

A deeper insight on the physical counterpart of a LGS is provided by the second Noether theorem, according to which the invariance of the Lagrangian under a group of local gauge transformations $\mathcal{G}$ implies that the currents which generate the corresponding global group $G$ are
the divergences of antisymmetric tensors

\[ J^a_\mu(x) = \partial^\nu G^a_{\nu\mu}(x) \quad G^a_{\mu\nu} = - G^a_{\nu\mu}. \quad (4.1) \]

(local Gauss law).

This is a very strong constraint on the physical consequences of \( G \) (corresponding to the Maxwell equations in the abelian case). Actually, such a property seems to catch the essential consequence of local gauge symmetry, since \( G \) invariance of the Lagrangian is destroyed by the gauge fixing, whereas the corresponding local Gauss laws (LGL) keep holding on the physical states, independently of the gauge fixing.\(^6\)

Moreover, a LGL implies that \( G \) invariant local operators are also \( G \) invariant. In the abelian case this implies the superselection of the electric charge ([13], Sect.5.3).

Thus, it is tempting to downgrade local gauge symmetry to a merely technical recipe for writing down Lagrangian functions, which automatically lead to LGL for the currents which generate the corresponding global gauge transformations.

The physical relevance of a LGL is that it encodes a general property largely independent of the specific Lagrangian model and in fact, most of the peculiar (welcome) features of Gauge QFT, with respect to standard QFT, may be shown to be direct consequences of the validity of LGL (see [13], Chapter 7):

a) a LGL law implies that states carrying a (corresponding) global gauge charge cannot be localized; this means that the presence of a charge in the space time region \( O \) can be detected by measuring observables localized in the (spacelike) causal complement \( O' \); this represents a very strong departure from standard QFT, where “charges” in \( O \) are not seen by the observables localized in \( O' \);

b) LGL provide direct explanations of the evasion of the Goldstone theorem by global gauge symmetry breaking (Higgs mechanism);

c) particles carrying a gauge charge (like the electron) cannot have a sharp mass (infraparticle phenomenon), so that they are not Wigner particles;

d) the non-locality of the “charged” fields, required by the Gauss law, opens the possibility of their failure of satisfying the cluster prop-

\(^6\)A gauge fixing which breaks the global group \( G \) involves a symmetry breaking order parameter and it is consistent only if \( G \) is broken (see [13], p. 178 and [15]).
erty with the possibility of a linearly raising potential, as displayed by the quark-antiquark interaction, otherwise precluded in standard QFT (where the cluster property follows from locality); e) a local gauge group may have a non-trivial topology, displayed by components disconnected from the identity, and the corresponding topological invariants defines elements of the center \( \mathcal{Z} \) of the local algebra of observables \( \mathcal{A} \); for Yang-Mills theories such elements \( \mathcal{T}_n(O) \), localized in \( O \), are labeled by the winding number \( n \) and define an abelian group \( (\mathcal{T}_n(O)\mathcal{T}_m(O) = \mathcal{T}_{n+m}(O)) \); their spectrum \( \{ e^{i2\pi n\theta}, \theta \in [0, \pi) \} \) labels the factorial representations of the local algebra of observables, the corresponding ground states being the \( \theta \)-vacua. They are unstable under the chiral transformations of the axial \( U(1)_A \) and therefore chiral transformations are inevitably broken in any factorial representation of \( \mathcal{A} \) without Goldstone bosons. Thus, the topology of \( \mathcal{G} \) provides an explanation of chiral symmetry breaking in QCD, without recourse to the instanton semiclassical approximation [13, Chap. 8].

In conclusion, LGS are not symmetries of nature in the sense that they reduce to the identity not only on the observables, but also on the states, possibly except for their local topological invariants. From the point of view of general philosophy, they appear in Gauge QFT as merely technical devices to ensure the validity of local Gauss laws (through a mathematical path which uses an invariant Lagrangian plus a non-invariant gauge fixing).

By the same reasons, i.e. the realization that the observables and the physical states are the only quantities needed for the complete description of a physical system, the claimed violation of determinism in gauge theories [1] is completely unjustified, since the observables and the physical states have a deterministic time evolution.

References

[1] K. Brading and E. Castellani eds., Symmetries in Physics: Philosophical Reflections, Cambridge Univ Press 2003

[2] F. Strocchi, An Introduction to the Mathematical Structure of Quantum Mechanics, 2nd ed. World Scientific 2008
[3] F. Strocchi, The physical principles of quantum mechanics. A critical review, Eur. Phys. J. Plus, 127: 12 (2012)

[4] F. Strocchi, Symmetry Breaking, 2005, 2nd ed. 2008, Springer

[5] F. Strocchi, Elements of quantum mechanics for infinite systems, World Scientific 1985

[6] D.M. Greenberger, Am. Jour. Phys. 46, 394 (2004)

[7] C. Liu, Philosophy of Science, 70, 1219 (2003)

[8] S. Doplicher, R. Haag and J.E. Roberts, Comm. Math. Phys. 13,1 (1969); 15, 173 (1969)

[9] P. M. A. Dirac, The principles of quantum mechanics, Oxford Univ. Press 1958

[10] K. Drühl, R. Haag and J.E. Roberts, Comm. Math. Phys. 18, 204 (1970)

[11] R. Haag, Local Quantum Physics, Springer 1996

[12] G. Morchio and F. Strocchi, Infrared problem, Higgs phenomenon and long range interactions, in Fundamental Problems of Gauge Field theory, A.S. Wightman and G. Velo eds., Plenum 1986

[13] F. Strocchi, An introduction to non-perturbative foundations of quantum field theory, Oxford Univ. Press 2013

[14] J. Frohlich, G. Morchio and F. Strocchi, Nucl. Phys. B 190 [FS3] 553 (1981)

[15] G. De Palma and F. Strocchi, Ann. Phys. 336 112-117 (2013)