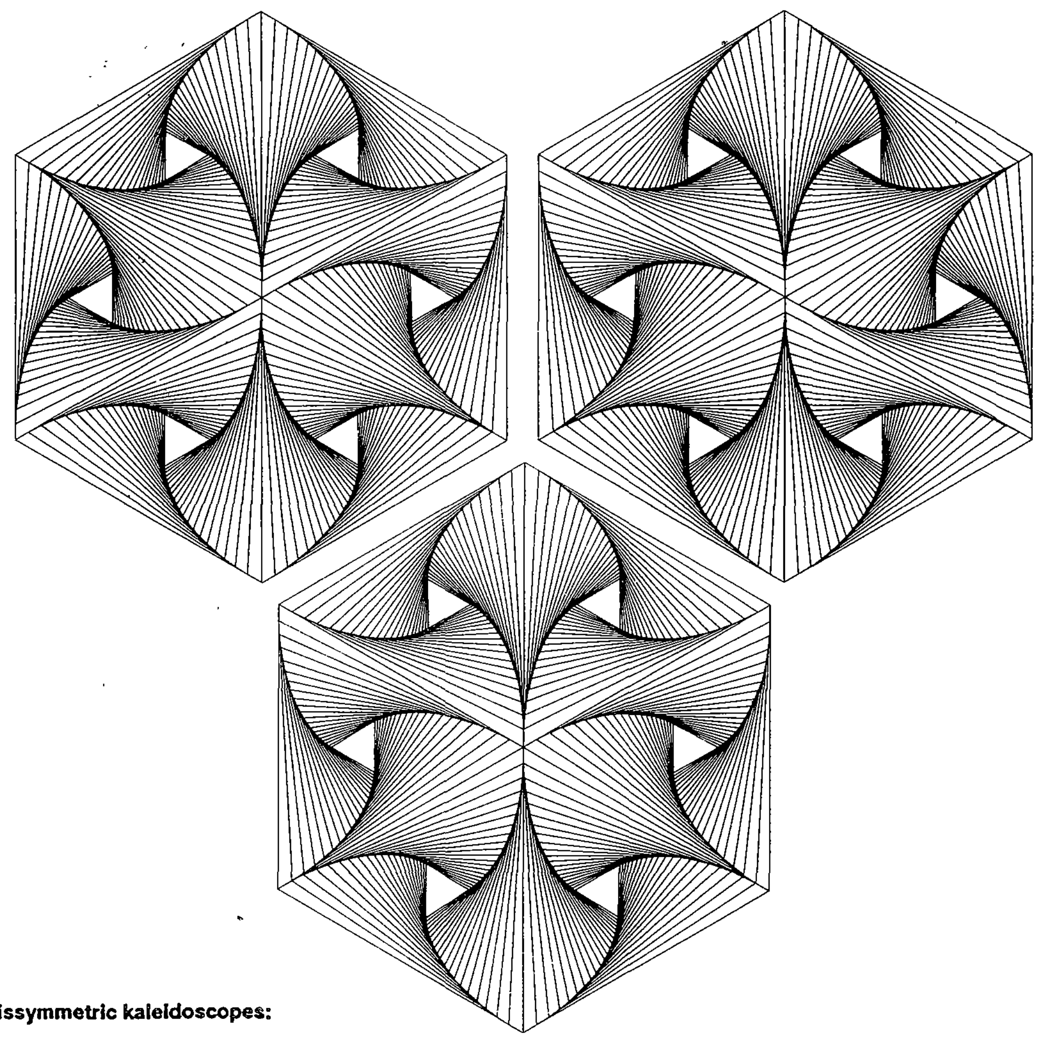
Symmetry: Culture and Science

SPECIAL ISSUE
Symmetry in a Kaleidoscope 3

The Quarterly of the
International Society for the
Interdisciplinary Study of Symmetry
(ISIS-Symmetry)

Editors:
György Darvas and Dénes Nagy

Volume 1, Number 3, 1990



Dissymmetric kaleidoscopes:
Hommage à Pasteur

Symmetry: Culture and Science
Vol. 1, No. 3, 1990, 229-255

SYMMETRY: A SPECIAL FOCUS ON ...

THE INTERPLAY OF SYMMETRY, ORDER AND INFORMATION IN PHYSICS AND THE IMPACT OF GAUGE SYMMETRY ON ALGEBRAIC TOPOLOGY

Yuval Ne'eman

Physics educator, government official, (b. Tel Aviv, 1925).
Address: Wolfson Chair Extraordinary in Theoretical Physics, Raymond and Beverley Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel 69978, *and* Center for Particle Theory, University of Texas, Austin, Texas 78712, U.S.A. (Supported in part by Grant DE-FGO5-85ER40200).
Fields of interest: Particle physics, statistical mechanics, group theory, cosmology, philosophy of science.
Awards: Weizmann Prize for Exact Sciences, 1966; Rothschild Prize, 1968; Israel Prize for Exact Sciences, 1969; Albert Einstein Medal and Award, 1970; Collège de France Medal, 1972; Wigner Medal, 1982.
Publications: *The Eightfold Way*, New York: W.A. Benjamin Pub., 1964 (with Gell-Mann, M.); *Algebraic Theory of Particle Physics*, New York: W.A. Benjamin Pub., 1967; *Symétries, jauges et variétés de groupes*, Montréal: Presses de l'Université de Montréal, 1979; *Group Theoretical Methods in Physics*, New York: American Institute of Physics, 1980 (with Horwitz, L.); *The Particle Hunters*, Cambridge: Cambridge University Press, 1986 (with Kirsh, Y.).

We review the role of symmetry in physics and of group theory. We explain the relationship between symmetries and conservation laws. In particular, the recent renewed geometrization drive is discussed, with gauge *local symmetries, sometimes spontaneously broken. Supersymmetry involves new generalized concepts. The interrelationship with order and with information is discussed. We survey the impact of the developments in symmetry in physics on the recent progress in several related fields of mathematics, especially in algebraic topology.*

CHAPTER 1: SYMMETRY – POSTULATES OF IMPOTENCE

1.1 In physics, the search for symmetry is a search for abstraction

Physics is an experimental and observational science and thus deals with the "real world". Its method, however, uses abstraction. The aim is to achieve a unified and coherent presentation of all natural phenomena. To treat different phenomena in a single formulation, physics has to strip away the circumstantial details and identify the essentials and discover the common denominators and their constrained behavior — the laws of physics. This then implies sweeping generalizations. The more phenomena are encompassed by a law — the more it has to become simple and rely on less specification.

Symmetry laws are in that category. They represent *negative statements embodying powerful generalizations*. They are "*postulates of impotence*" as stated by Whittaker in 1949, though highly potent ones. *Impotence*, because they state that it is impossible to prefer one frame over the rest. If a crystal is hexagonal, it has a symmetry under rotations by 360°/6 = 60°, and it is impossible to select one face out of the six as a "preferred" face. The *principle of covariance* in Einstein's general theory of relativity (the classical, i.e. macroscopic, theory of gravity) states that it is impossible to select a preferred reference frame — i.e. the laws of gravity do not depend on the selection of a particular reference frame, all reference frames are equivalent. The French saying goes "la nuit, tous les chats sont gris" — at night, all cats are grey — i.e. it is impossible to distinguish or specify a preferred cat. There is then a symmetry between cats, they all look the same.

1.2 Active and passive transformations

Symmetry can be *passive* or *active*. For example, Einstein's principle of covariance, one of the two pillars of the general theory of relativity, is a *passive* symmetry, since it proclaims the inexistence of a preferred reference frame, i.e. a preferred *coordinate system*. We compare two Cartesian systems of axes in the plane, for instance: one with the x-axis lying horizontally, and one in which it is at 45° (or at 01.30 hours). We do not rotate the physical system at study, only the coordinates.

In an *active* symmetry transformation, we keep the same coordinates, but we *really move* the physical system (or modify it, in the case of *internal* qualitative symmetries that are not related to motion). When discussing for instance the earth's motion around the sun, we compare the state of the sun-earth system in winter and in summer, at different phases of the rotation. In a passive symmetry discussion we would have kept everything in place and only rotated the coordinate system.

Most symmetries can be used in either manner, but not all — as in the example of covariance.

CHAPTER 2: ALGEBRAIC FORMALISM

2.1 Group theory

The basic mathematical tool that is used in dealing with symmetry is the theory of groups. It was invented by a twenty-one year old French student, Evariste Galois (1811-1832), and written up as a "mathematical testament" in one night, prior to his responding to a duel challenge in the morning — a duel which did cost him his life.

A *group* is a set of *elements*, any two of which can be combined to make a third element. For instance, the real integers (negative and positive) — i.e. the set ... -3, -2, -1, 0, 1, 2, 3, ... with addition (+) as the *combining operation* — fit this description. In addition, there should be a *neutral* element, such that if combined with another element a, the resulting element is still a; in our example, zero fulfills this role, since $a + 0$ is still a. The last requirement is that there should be an *inverse* element, i.e. an element $1/a$ which, when combined with a gives the neutral element. In our example, -3 is the inverse element for 3, because $-3 + 3 = 0$, etc.

A clock can serve as another example for a group. The elements are time-intervals (given in minutes and in seconds). We set the clock on XII or zero hours; we now move the hands so as to add some time lag and then continue with a second time interval. Combination here amounts to adding angles cyclically. In other words, the result of combining the elements

"5 minutes and 10 seconds" with "3 minutes"

is the same as the result of combining "6 hours and 4 minutes" with "6 hours 4 minutes and 10 seconds". In this example, the inverse of "3 hours" is "9 hours" because their combination will yield XII or "0", the neutral element.

This second example belongs to a class of groups known as *transformation groups*. In such groups there is a substratum to which one applies transformations, changes (here it is the clock, i.e. a circle with cyclic coordinates along it). The transformations make up the group-elements; the combining operation consists in applying the two transformations one after the other.

A group can be either *discrete* or *continuous*. Another name for such continuous groups is *Lie groups*, after a Norwegian mathematician who first studied them, Sophus Lie (1842-1899).

In our second example, we had a continuous group. But we can also apply a discrete *subgroup* of that group: suppose we limit the elements to 60° angles, i.e. any even number of hours. Two hours make a 60° angle clockwise; combining "-4 hours" with "22 hours" yields "18 hours", which coincides with "6 hours". There will be only 6 distinguishable group elements in that group. If we always replace the number of hours by its equivalent in the interval 0 — 10 (since 12 ~ 0) we have a *finite* group with just 6 elements. This is the symmetry group of a hexagonal crystal or of a snowflake. If we keep all even hour values such as "-4" or "22" or even "50" (without replacing "50" by "50"–"48" = "2", although we situate "50" at the same place as "2") we are applying the *covering group*, in this case an *infinite (countable) covering group*.

2.2 The rotation groups and non-commutativity

One of the most utilized groups in physics is a continuous group, the group of *transformations on the sphere*, also known as the *rotation group*. We take a globe or a ball and we apply to it various rotations. We combine these elements by applying two transformations consecutively. The neutral element is known as the *identity transformation* because we do nothing, so that the position of the globe after this transformation is identical with the position it was in before applying that transformation.

This group is the *rotation group in 3 dimensions* — the 3 dimensions of the space in which the sphere exists and rotates. It could not be fitted in a flat 2-dimensional space such as a flat piece of cardboard — or our clock's dial. Indeed, our clock group is the *rotation group in 2 dimensions*.

There is a very important difference between these two rotation groups. In the 2-dimensional group (the clock) the order between two different elements does not influence the result of their combination: "2" + "4" = "6" and "4" + "2" = "6", too. This is a *commutative group*, also known as an *Abelian group*, named after another young mathematician, the Norwegian Niels Henrik Abel (1802-1829) who died of tuberculosis.

The 3-dimensional rotation group is *non-commutative*. You can make an experiment: take a book, put it on the table or desk in front of you. Denote two sides of the front cover by x and y (x is the side that is parallel to your abdomen, y is perpendicular to it). z is the vertical axis, perpendicular to the front cover (and parallel to the book's "thickness" dimension). These axes should be regarded as fixed to the book cover, moving with it.

Now rotate the book by 90° clockwise (when looking along the relevant axis from its origin, left to right) twice: first around the x axis and then around the y axis. The book ends up standing with its spine to you. Now put it back in the original position and try the same thing, but in the opposite order: first around y, then around x. The book ends up standing on its long side, with the front-cover facing you.

We can try the same exercise with the x, y, and z axes defined with respect to the room. You will get the same results except that they will be inverted: x followed by y will yield the result we got previously from y followed by x, and vice versa. In any case, the order matters, so that the group is clearly non commutative.

2.3 Parity and *CP*

In physics, we describe the symmetry by the group of transformations to which it corresponds. One such symmetry is *parity*, corresponding to the 2-element group of reflections in a mirror. Electromagnetism, gravity and the *strong* nuclear force that glues quarks within a proton, or protons and neutrons within an atomic nucleus — all these forces obey this symmetry (denoted by P).

And yet it was found in 1956 that another *nuclear* interaction, the force that causes, for instance, neutrons to disintegrate (*beta decay*) violates parity. When the particle known as the *mu-meson* or *muon* (a *heavy electron* with a mass 207 times that of the electron) disintegrates, it yields an electron, a *mu-neutrino* and an *electron-antineutrino*. It turns out that neutrinos always spin like left-handed screws, whereas antineutrinos screw rightwise. Had parity been obeyed, we would have found either type of particle screwing in either direction.

It is interesting that these forces — the *weak interactions* sometimes named after the Italian physicist Enrico Fermi (1901-1954) — do obey a different *reflection* *group*: in addition to spatial mirror reflection, also *invert all charges* — electric and other, such as *leptonic* (carried by a neutrino, for instance). Thus you have to replace an electron by a positron, a neutrino by an antineutrino, etc. This charge *conjugation* (or inversion) is denoted by C and the combined inversion by PC (or CP). In 1964 it was found that a few interactions even violate CP invariance.

In the case of the rotation groups or parity, the transformations occur in physical spacetime configurations. If, however, we discuss the CP transformations as a reflection, we are going beyond spacetime: the inversion of charges occurs in an abstract *charge-space* that has somehow become soldered here to ordinary space.

In the physics of particles and fields there are many such *charges* and our transformations generally occur in their abstract spaces, which have come under the term *isospaces*. There are symmetries in which we have a 3-dimensional isospace and the group is the rotation group as in its action on a physical sphere — except that the sphere is an abstract object that has nothing to do with spacetime. Such is the symmetry between protons and neutrons: these particles behave as if they were *spinning* in opposite directions — but in this mathematical isospace, describing *internal*, i.e. qualitative, features that do not reflect on spacetime.

2.4 Unitary symmetry ($SU(3)$)

One such symmetry is an abstract invariance of the strong nuclear "glue" under the Lie group $SU(3)$ — the *Special Unitary group in 3 complex dimensions*. Imagine a 3-dimensional *complex* space — here it consists of three generalized charges (also named *flavours*) somewhat like the electric charge. Their complex nature means that they are measured in complex numbers, such as $5 + 3i$, where $i=\sqrt{-1}$ is the imaginary unit, the square root of -1. Under the charge-conjugation operation C, they go over to the complex-conjugate quantities, $5 - 3i$ in the above example. Electric charge is also described by *complex quantities*, with complex conjugation taking us from a positively charged state to a negatively charged one. Two *real* quantities are thus involved: in the above example, these are the 5 (the real part of the complex quantity) and the 3 (the real coefficient of $\sqrt{-1}$, i.e. of the imaginary part) in electric charge; however, rather than their appearing as separate entities, they are intertwined within one complex quantity and its conjugate.

For $SU(3)$, this means that there are in fact 6 *real* charges, appearing in the three complex combinations and their conjugates. The simplest system capable of carrying 3 such complex charges (non-trivially, i.e. other than zero amounts) corresponds to 3 *particles* or fields. No such object had been observed when the symmetry was discovered and they were hypothesized as either real states or mathematical toy models to understand the symmetry (Goldberg and Ne'eman, 1963; Gell-Mann, 1964; Zweig, 1965).

The group $SU(3)$ consists of all possible transformations that preserve the *lengths* of complex vectors, just as the rotations also preserve lengths in a real space. Length preserving transformations are called *unitary* — this is the U in $SU(3)$.

In addition, these transformations also preserve the equivalent of a volume in the complex context; this is what the word *special*, the S in $SU(3)$ stands for. There are nine different possible transformations between 3 *flavours* u, d, s: $u \to u$, $u \to d$,

$u \to s$, $d \to u$, $d \to d$, $d \to s$, $s \to u$, $s \to d$, $s \to s$ (the names relate to "up", "down" and "sideways", or "singlet" originally). They can be combined in many ways, but one combination amounts to doing nothing: $u \to u$, $d \to d$, $s \to s$. This leaves eight basic transformations, which is why the colloquial term for $SU(3)$ is the *Eightfold Way*, a name introduced by one of the discoverers of that symmetry in nuclear particles (Gell-Mann, 1961 [unpublished]; 1962). The scientific term is *unitary symmetry*, generally used by the other discoverer (Ne'eman, 1961). (The symmetry was discovered by both physicists simultaneously and independently).

2.5 Representations

A group can have many different representations. The simplest is the *trivial* representation, i.e. the invariant or the scalar. For instance, temperature is invariant under the application of rotations in our room. We apply a rotation, and yet nothing happens as far as temperature is concerned. Another example is the area within any closed curve drawn on a sheet of paper, when acting with 2-dimensional rotations. If the figure is a rectangle with its longer side drawn horizontally and we rotate the drawing by 90°, the sides will have all changed their configuration (they are *vectors*), but the area has remained unmodified.

For rotations in 3-dimensions we can have vectors and tensors of any order, i.e. more complicated representations.

We can take an example in $SU(3)$. The *basic* representation that was used to define the group involved a complex 3-dimensional *vector*. The fields that actually carry the 3 flavour charges are known as *quarks*. Introduced as either true particles or a toy model, later (1964-69) experimental and theoretical results showed that they do not exist as separate particles, but that they do have a physical existence when "confined" within protons, neutrons, hyperons, and mesons. One can sense their presence when probing inside such particles with extremely hard gamma rays.

In any case, the quarks span the 3-dimensional representation of $SU(3)$. Denoting them as before by u, d, s, the anti-quarks u^*, d^*, s^* carry the conjugate (or inverted) charges and span the 3 complex dimensions of the conjugate representation 3^*.

Let us now try to list all possible combinations of 3 such quarks: *uuu, uud*, etc. We can arrange them in an array

```
ddd  ←   udd  ←   uud  ←   uuu
 .      /   .    /    .    /
  .   /      . /       . /
    dds  ←   uds  ←   uus
      .      /   .    /
       .   /      . /
         dss  ←   uss
           .      /
            .   /
             sss
```

In this array, moving one step along the arrows transforms a u quark into a d quark. Moving parallel to the broken line down and to the left, transforms in each step a u quark into an s. Moving along the dots, down and to the right, transforms a d quark into an s. This array is thus a 10-dimensional (you can count the various 3-quark states) representation of $SU(3)$. In fact, it is an important representation historically, because such 3-quark states were known in 1962, except for the sss combination and one could therefore predict the existence of that particle state with all its specific properties. It was indeed discovered experimentally in 1964. There are nowadays entire beams of *omega-minus* particles (the sss combinations). For a phenomenological and historical treatment of particle physics, see Ne'eman and Kirsh (1986).

The above example illustrates an important feature in the applications of symmetries at the particle or field level: they provide a *classification*. Particle states or fields have to form multiplets of the relevant symmetry group. In many cases in particle physics the sequence is even inverted: it is through the identification of multiplets that one guesses at the symmetry.

2.6 Lie algebras and conservation laws

A very important role is played by infinitesimal transformations. Readers who are not familiar with the calculus should think of very gradual transformations, in extremely small steps. Suppose we apply a rotation to this page. The amount we rotate by is an angle. We could make a 30° rotation in 30 small steps of 1° at a time. The angle of 30° is the magnitude of the *parameter* of this Lie transformation. The rotation per degree is called the *algebraic generator*. It represents the small (infinitesimal) transformation. If repeated 30 times it will yield the finite rotation we wished to perform — except that it should be applied like a compound interest: each new tiny application acts on the system by taking it from the position that was reached through the combined effect of all previous applications (mathematically, this is an exponentiation).

The algebraic generator of rotations is called the *angular momentum* and denoted by J. In 3 dimensions there are 3 independent ones: $x \rightarrow y$ (or $y \rightarrow x$, which is the same rotation), $y \rightarrow z$ and $z \rightarrow x$. They are denoted J^1, J^2, J^3.

There is an algebraic generator for each dimension of the group: in $SU(3)$ for each of the eight possible types of transformations that we have enumerated. The 8 algebraic generators of $SU(3)$ are called the *unitary spin* and are denoted by F (F^1, F^2, etc.).

The study of the role of symmetry and invariance goes back to the late eighteenth century (Euler, Lagrange) and the nineteenth (Jacobi, Hamilton, F. Klein). It was further advanced by an important theorem due to Emmy Noether (1918). The theorem was derived for classical physics, but it has acquired much more importance after the advent of quantum mechanics.

The dynamical time-evolution of a system is represented in physics by the *Hamiltonian* (a function in classical physics, a mathematical operator in quantum physics). Let us think of the Hamiltonian as the algebraic generator H for time-displacements. It will represent the time-displacement per tiny unit of time — say a *picosecond* (one "American" trillionth of a second) just as the angular momentum generators J^1 etc. were defined as the rotations per degree.

We can now write statements,

$$[H, J^1] = 0, \; [H, J^2] = 0, \dots, \; [H, F^5] = 0, \; \text{etc.}$$

The symbol [,] is a *commutator*. It represents the difference between acting (on a system on its right) first with J^1 and then with H and acting in the inverse order. A vanishing commutator means that in acting on a system, the order between H and J^1, for instance, thus does not matter.

This can be understood in two ways.

In one interpretation, we think of H acting on the system, i.e. displacing it in time, aging it. The statement then says that applying J^1, for instance, onto H has made no difference: "the dynamical history of the system is unchanged by a rotation". It is a *message of symmetry*; it is a short-hand way of saying that the physical laws are spherically symmetric, invariant under rotations (that is so if the commutator is also vanishing with J^2 and J^3).

But there is another conclusion. Let us think of the Hamiltonian H as the time displacement generator. If the application of J^1 or F^5 to the system is not modified by a time displacement, this means that J^1 or F^5 represent *conserved quantities* or generalized conserved charges.

This is the essence of the theorem of Emmy Noether: for every continuous symmetry there is a conservation law and the conserved charge is identical with the algebraic generator of the symmetry — and vice versa.

Thus $SU(3)$ corresponds to the conservation of the F^1 to F^8 charges of unitary spin; angular momentum J^1 to J^3 is another such set etc. Taken as a charge, H itself represents energy conservation. The conservation of linear momentum P^1 to P^3 is the result of symmetry under spatial displacements.

CHAPTER 3: BROKEN SYMMETRIES

3.1 Laws of nature versus boundary conditions

Symmetry can sometimes be in the *laws of physics* and then has a great range of applications: in the example of Einstein's theory of gravity, for instance, whatever the gravitational problem, the laws will still have to be stated *covariantly* (i.e. independently of the selection of a reference frame, of a coordinate system).

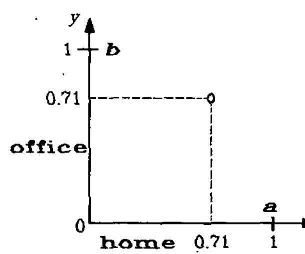
Sometimes, however, there is a symmetry that relates to the *boundary conditions*. In the Kepler problem (sun and planets) for instance, there is an *a priori* spherical symmetry in the given conditions themselves: the sun is assumed to be spherical, and therefore there will be no preferred direction for the its gravitational pull — in the way that would happen in a description of gravity in our room. It so happens that the laws themselves also contain no preferred direction and are spherically symmetric, even for our room, in which the boundary conditions are less symmetric.

In fact, the symmetry of the laws is generally greater than that of the given ones; in the case of Einstein's theory, for instance, the laws are also locally Lorentz-invariant, which includes, aside from insensitivity to rotations of the system in space, an invariance under accelerating boosts (*special Lorentz transformations*).

Sometimes, we are surprised by the amount of symmetry sustained by the boundary conditions. In cosmology, for example, there is no known *a priori* reason for the boundary conditions to be very symmetric. They could have been as complicated and asymmetric as we wish — and yet in reality, the observations show that the cosmological boundary conditions are highly spherically symmetric.

3.2 An asymmetric vacuum state

In modern treatments, there is a delicate interplay between laws and boundary conditions. We shall see that symmetry has to be broken at some stage, when we deal with the real world. In the words of Francis Bacon, "there is no excellent beauty that hath not some strangeness in the proportion". Rather than break the symmetry of the laws, it is more convenient — and useful — to find formulations in which the laws are entirely symmetric, and the symmetry breakdown is "blamed" on some boundary conditions. In quantum mechanics, the "real world" is given by the Hilbert space, the abstract space in which each *dimension* represents one position in spacetime. For instance, if I can be either at home or at the office, we can draw a 2-dimensional diagram



in which the abscissa is related to the probability that I am at home and the ordinate relates to the office. The probability of my being in either place is given by the square of the length of the projection on the relevant axis. When I am sure to be home, my state is described by the point a: $x = 1$, $y = 0$, so that $x^2 = 1$ and also $x^2 + y^2 = 1$. When I am sure to be at the office, the state is described by the point b: $x = 0$, $y = 1$, so that $y^2 = 1$ and again $x^2 + y^2 = 1$. The point o represents a state in which $x = 0.71$, $y = 0.71$; 0.71 is roughly $\sqrt{2}/2$, so that $(0.71)^2 = 1/2$. Thus, $x^2 = 1/2$, $y^2 = 1/2$ and there is a 50% chance for my being at home and another 50% probability for my being at the office. The total of the probabilities has to be 100%, and indeed $x^2 + y^2 = 1$ again.

If a particle can be at 29 different places, the Hilbert space will have to be 29-dimensional. Most situations correspond to particles that can be at an infinite number of positions, and in most cases the Hilbert space has to be infinite-dimensional.

We now return to the breaking of symmetry. In most cases this can be understood in the following way: the laws do obey the full symmetry, but the basic state in the Hilbert space, the *vacuum state* does have a preferred direction.

The vacuum state is the state in which there is *nothing* (although in quantum mechanics you can extract a lot from the vacuum for a brief time, within the *error-brackets* provided by the Principle of Uncertainty). To get a one-particle state, we have ways of *constructing* the particle onto the vacuum state. We can see to it that the construction should not modify the basic direction relating to the characteristics of symmetry: if, for instance we are dealing with a type of *charge* (that is not explicitly conserved because the symmetry is broken) we already endow the vacuum with a certain amount of that charge, and the particles built on this vacuum will also have that feature. In this manner, we continue to have a preferred direction imposed by the boundary conditions of the problem, in this case the Hilbert space.

3.3 Superconductivity as a model

This approach was first introduced in the study of superconductivity from the physics of condensed matter. In that discipline, this method was invented (Ginzburg and Landau, 1950) to explain phase transitions, such as the transition in a material between a paramagnetic and a ferromagnetic state when it is cooled down to the critical temperature – or the transition to the superconducting state at very low temperatures (since 1985, the temperatures are no more that low). In a more structural theory of superconductivity (Bardeen, Cooper, and Schrieffer, 1957) we can understand the asymmetric behavior of the vacuum.

In that problem, a *false vacuum* state is created, when the overall interaction between the electrons and the atomic lattice in the metal produces a *pairing* between electrons: two non-contiguous electrons start acting as if they were bound. This then becomes the lowest-energy ground state and acts as a vacuum for that particular situation; but this vacuum is not really a neutral empty vacuum, and thus contains characteristics that break the symmetry of the equations.

This idea has been described as a *spontaneous* breakdown of the symmetry. The method was successfully generalized to the physics of particles and fields (Nambu, 1960; Nambu and Jona-Lasinio, 1961; Goldstone, 1961). Here, the assumption of a *directed* vacuum requires the existence of massless particles – massless in the approximation in which all other effects are removed. The massless particles are needed to complete the vacuum's multiplet.

In an unbroken symmetry, the vacuum is invariant, i.e. if we apply to it the symmetry transformations, it does not change. In other words, the symmetric vacuum is a *scalar*, forming a single-state multiplet.

But when the vacuum has a direction, applying the symmetry operations to that state should rotate it into some other state. What would that state be like? It turns out that a particle with zero mass could serve as a *partner* for our non-single vacuum.

3.4 Chiral unitary symmetry

The idea was very successfully applied to the understanding of the Yukawa force. This is the force responsible for the attraction between nucleons (protons and neutrons) in any atomic nucleus. It involves the exchange of pions (the *meson* postulated by Yukawa and Stueckelberg in 1934) between nucleons, like volley balls in that game. The force obeys a certain symmetry called $SU(3) \times SU(3)$ *-chiral*, because the relevant conserved currents are characterized — on top of the *unitary-symmetry* charges they carry — by left or right *handedness*. The two $SU(3)$ in the name of the symmetry correspond to two currents, one an $SU(3)$-left and the other an $SU(3)$-right. Note that parity is conserved because both chiralities are present; it is only when the left-chiral current of $SU(3)$-left comes by itself — in Fermi's weak interaction — that parity is thereby broken.

The doubling of the $SU(3)$ currents and symmetry is quite analogous to what we observe in the case of angular momentum. In very low energy atomic physics we can have a separate conservation of spin and orbital angular momentum, i.e. two $SU(2)$ currents of angular momentum. However, once we increase the energies involved, the spin and orbital angular momenta mix, and only total angular momentum is conserved. The same happens with the unitary symmetry chiral currents. Once the symmetry is broken, only the sum of $SU(3)$-left + $SU(3)$-right subsists as a conserved quantity. This sum is plain $SU(3)$, and in a certain approximation it is even *locally* conserved. Its currents then couple *universally* to an octet of spin 1 vector-mesons.

Chiral unitary symmetry together with this $SU(3)$ *gauge* provide a good phenomenological working theory for the physics of hadrons — the hundreds of different particles that feel the *strong* nuclear interaction and that we now consider as consisting of bound systems of either three quarks or a quark and an antiquark. The theory is sometimes described as *current algebra*. It fuses two theoretical discoveries of 1959-64: unitary symmetry $SU(3)$ (Ne'eman, 1961; Gell-Mann, 1961) and spontaneous symmetry breakdown (Nambu, 1960; Nambu and Jona-Lasinio, 1961; Goldstone, 1961) using techniques (Gell-Mann, 1962) inspired by Heisenberg's version of quantum mechanics, the *matrix mechanics* (Ne'eman, 1967; Adler and Dashen, 1968).

CHAPTER 4: LOCAL (GAUGE) SYMMETRIES

4.1 A financial interlude

The global symmetries we discussed in Chapter 2 are mostly useful in the first *exploratory* research phase in the physics of particles and fields: they provide a classification, which then leads to the understanding of structure. The experience of the last 75 years has been that the *dynamical* theories are based on *local* symmetry principles.

The first such symmetry was *general covariance*, a passive symmetry principle of general relativity. The other principle of that theory, the *principle of equivalence*, is to a certain extent the model of how a local gauge symmetry should be actively implemented. This is in fact the meaning of its name. We shall return to this question in a later part of this chapter. First we shall present the generic local gauge principle, the Yang-Mills gauge (Yang and Mills, 1954).

We start with an analogy. Let me assume (for a few brief minutes) that I am a wealthy businessman, deeply involved in international deals. Almost daily, and sometimes more than once a day, I have to transfer sums of money from country to country. In each country, there is a *local* interest rate that is fixed by the local government. As a result, I have to be very careful and think twice before ordering my bank in country A to transfer funds to country B. Indeed, if the interest rate in country B is much lower than in country A where my money is now, and should the contemplated transaction be delayed or cancelled, I shall have suffered an unnecessary loss.

I can resolve this difficulty and enjoy an effectively fixed interest rate by making a deal with an international banking concern. I shall deposit my money with them and make all my transfers through their branches in the various countries, provided they undertake to compensate me for any loss incurred because of the variations in interest rates. They cannot just give me a fixed rate internationally (a *global symmetry)* because of the local legal implications: they might loose their permit to operate a bank in country C, should they disregard the legal rate in that country.

What the international banking corporation can do is to establish an auxiliary corporation (called GF for General Financing — or for *Gauge Field*) that would own and run in each country some business operation out of which they will compensate me for the difference between the local interest rate and some standard that we fix together (perhaps the rate in the USA, for instance). The compensation will have to be indirect: they could, for instance, sell me shares in some firm owned by GF and offer me a reduced price — precisely by the same amount that I lost over the lower local interest rate.

This method appears much more complicated than having a unique *global* rate, but it might still represent the only viable solution for this financial problem. The implication is that aside from any other business I might be involved with in country B, I shall also be interacting everywhere with the GF corporation. It will be through this additional interaction that the compensation will be able to take place and the symmetry between all my financial ventures — whatever the country — will be restored.

4.2 The Yang-Mills gauge

It is time to wake out of my financial dream — for a few minutes I could almost believe I would end up making it in *Fortune* magazine.

Yet I do believe that physics is no less exciting. Imagine an *internal* symmetry — i.e. a symmetry that does not act on the spacetime coordinates. Assume the relevant symmetry parameters, such as the amount by which we rotate a state in its isospace orientation — changes from place to place. We allow an arbitrary selection of angles at each place. This is quite natural, since we cannot communicate with all other places instantaneously and it would thus be impossible to organize arbitrary global active transformations without preparation.

In an example such as flavour-$SU(3)$ such a transformation consists in replacing, e.g. a u quark by a d quark — in practice a proton ($= uud$) by a neutron ($= udd$). With the superposition principle of quantum mechanics, it can even consist of gradual transformations of this type. However, while we are changing $u \to d$ at x, somebody is changing $d \to s$ at y. The only way in which these changes will not introduce inconsistencies is through the existence of a *compensation* system GF: the gauge field.

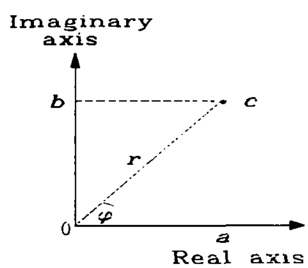
The gauge field will interact with the charge-currents of our symmetry — they will represent the sources that induce its presence, in the corresponding equations. The interaction between the charge-carrying particles or fields and the gauge field will result in a restoration of a symmetry — a very large one in this situation. It is the inexistence of preferred reference frames in isospace (this is the postulate of impotence for a global internal transformation) multiplied infinitely because we can now select any reference frame out of all the possible ones at x and have that freedom independently at all the infinite possible positions x in spacetime.

The electromagnetic field is such a gauge field, for the Abelian (commutative) group $U(1)$ of rotations in the complex plane or phase transformations.

Remembering the fact that a charged field is described by complex numbers c (and that the inverse charge is given by its complex conjugate c^*), we replace the representation of the complex number $a + bi$ (where $i=\sqrt{-1}$ is the imaginary unit) by its representation by a modulus r (the length of the vector from the origin in the Argand diagram for the complex plane) and a phase φ (the angle between the real axis and the vector from the origin),

$$c = a + bi = r\exp(i\varphi)$$
$$c^* = a - bi = r\exp(-i\varphi)$$

where $r^2 = a^2 + b^2$, $\tan\varphi = b/a$



and the Lie group is the rotation group in the complex plane $U(1)$, so that applying $\exp(i\alpha)$ to the above c, this will rotate it to $c' = r\exp i(\varphi + \alpha)$. Here we are using $\alpha(x)$, i.e. we are changing φ by locally dependent amounts. This would have been impossible — it would have introduced arbitrary gauge-dependent quantities such as α and $\partial\alpha$ into the physical equations — except for the presence of the electromagnetic field γ. The variation of the electromagnetic field itself is also locally dependent (it will involve $\partial\alpha$) and it is adjusted so as to cancel the arbitrariness introduced by the local freedom in selecting the transformations. *The gauge field does it by interacting with the electromagnetic charge and current densities.*

The electromagnetic charge and current (the flow of the charge) are the sources of the electromagnetic field in Maxwell's equations. Forty years after Maxwell introduced his equations, their global symmetry, the Lorentz group which dominates the world of the special theory of relativity, was understood by Lorentz, Poincaré, Einstein, and Minkowski. Twenty-three years later (Fock, 1927; London, 1927; Weyl, 1929), the local symmetry of the quantum version of Maxwell's equations was also understood, following an unsuccessful first try by Hermann Weyl who tried to identify that gauged rotation with a non-quantum feature, a scale transformation (Weyl, 1918).

The same type of local symmetry with a corresponding gauge field can be generated by any Lie group applied *internally* (i.e. a group that does not act on the spacetime manifold itself) — (Yang and Mills, 1954; Shaw, 1955). For a non-Abelian (i.e. non-commutative, see 2.2) group, the gauge field itself has more than one component; it should have one component per dimension of the group space, or per independent parameter.

The gauge field is thus itself a multiplet of the group, the *regular* or *adjoint* representation, the same representation as that of the algebraic generators of the group, the charges with which the field components will interact.

Since *the Yang-Mills field itself* is a non-trivial multiplet, it *also carries the group charges*, like other representations. It will therefore contribute to the charge-current density and will thus interact with itself. This is not so with the electromagnetic field, which does not carry electric charge. As a result, the spacetime dependence of the resulting force is different too: as against the $1/r^2$ dependence of the Coulomb force; we have here a constant, range-independent force.

4.3 Quantum chromodynamics and confinement

We have learned an interesting and unexpected lesson in the seventies. It turns out that *all four forces that we understand are gauge interactions*! One is classical general relativity to which we shall soon return for another look at the principle of equivalence. We do not know what quantum gravity will be like — it is one of the important open problems in physics — but it will be surprising if it does not have even more local symmetry.

The fundamental nuclear glue is a gauge interaction with $SU(3)$ as the local gauge group. Note, however, that this is a different $SU(3)$, known as $SU(3)$-*colour* (Han and Nambu, 1965; Fritzsch and Gell-Mann, 1972) as against the "older" $SU(3)$-*flavour.*

$SU(3)$-*flavour* — or even its extension as a 9-parameter group $U(3)$ in which the conservation of *baryon charge* (or *atomic weight number*) becomes an integral part of the symmetry — also *has an effective gauge interaction* which dominates the region of energies between 1 – 1000 GeV, with mesons known as the $\rho^{\mp}$, ρ°, $K^{\mp}$,

K°, $\overline{K}^\circ$, ω°, Φ° as the gauge fields. These mesons have the correct spin 1 like the electromagnetic field γ°. However, as the symmetry is only approximate, they are massive and short ranged accordingly. In practice, we know that they are also quark-antiquark compounds.

Colour-SU(3) is a precise symmetry, therefore probably more fundamental. Its 8 vector mesons (*gluons* Γ) are massless and would have been long-ranged like γ° if it were not for the confining property (Fritzsch and Gell-Mann, 1972).

This feature is a consequence ('t Hooft, 1972; Gross and Wilczek, 1973; Politzer, 1973) of the constant strength of the force — upon condition that the number of matter fields does not increase beyond 16 flavours (today, we would say 8 *generations*). It causes any particles carrying the colour-$SU(3)$ charges to be confined within systems whose total colour-$SU(3)$ charge vanishes: either quark-antiquark combinations in which the colour cancels mutually, or 3 quark combinations in which there are always mixtures of the 3 colours in which the total vanishes — just as we get a white colour from mixing red, blue, and yellow.

The Yang-Mills interaction of colour-SU(3) is called QCD (Quantum Chromodynamics). The mesons that mediate *chiral* flavour $SU(3)$ symmetry (through the Nambu-Goldstone mechanism that we reviewed) are in fact quark-antiquark compounds glued by the QCD gluons — even though they manage to fulfill an approximate dynamical and symmetrical role in addition, like the $SU(3)$-flavour vector mesons.

4.4 Gravity as a local gauge

There are two important features — universality and equivalence — that are specific to the dynamics deriving from local gauge theories.

Both features are already present in Einstein's general theory of relativity, the only dynamical theory which already at the classical level displayed several of the features of a local gauge theory. Remember that the gauge nature of Maxwell's theory only appeared at the quantum level, with the gauged phase as group parameter.

The group that is *locally gauged* in the theory of gravity in the full sense (i.e. both passively and actively — we shall return to this point in what follows) is the *Poincaré group*, with some adaptations that we shall discuss. The Poincaré group includes the Lorentz subgroup (rotations and Lorentz boosts — i.e. changes of velocities) together with translations in space and time.

Only the Lorentz subgroup in its *spin* action, — i.e. when it acts intrinsically and not orbitally — really parallels the (simpler) Yang-Mills model; the translations and orbital rotations or boosts introduce complications because they change the point in spacetime at which the transformation was defined. Remember that the Yang-Mills gauge involves arbitrarily different transformations at different points, with the gauge field supplying parallelism. The transformations at the position x stayed at x. Here, when we perform an x-dependent translation *in space* itself (rather than in an isospace as in Yang-Mills) at x, we end up arriving at y! Our problem is that the group parameter, the amount by which we transform, is a length in spacetime itself.

This is resolved through the utilization of *a special mix of a passive with an active transformation*: first we perform a *passive* transformation, just changing the coordinate's origin and inclinations so that the point in space that was known as x is now given by y. Nothing but a change of name — which is why this passive transformation is called an *alias*. Then we perform an *active* transformation, moving the system (a matter field, for instance) from its position (formerly called x and presently called y) to a new position whose coordinates in the new system of coordinates will have the value x. This is a real change of place, which is why the transformation is nicknamed an *alibi* transformation. In any case, the result is indeed — formally at least — a *local* transformation, since we start and end at x. We could also invert the order and perform the alibi first (move really from x to y) and then have an alias transformation that would rename y as x.

4.5 The principle of universality

Let us first discuss the principle of universality. In Newton's theory, the strength of the coupling of matter to the gravitational field is given by the mass of that matter; in Einstein's theory this feature takes on a relativistic profile and the *charge* and *current* that couple to gravity are given by the various components of the energy-momentum tensor. In fact, the *static* Newtonian component is given by the *energy*, and for a body at rest, this is Mc^2, i.e. it is again the mass (up to constants that are incorporated in the units) and we are back with Newton's third law. On the other hand, for a photon of light, it is indeed the energy $E = h\nu$ according to Planck's hypothesis (the photon has no mass) and this gives in the general theory of relativity the correct observed deflection (e.g. of rays passing close to the Sun, as observed during the 1919 eclipse).

When we follow Emmy Noether's theorem and derive the conservation laws resulting from invariance under the Poincaré group we find that the conserved charge-current-density corresponds to the ten components of the energy-momentum tensor. The static conserved *charge* is indeed just the *energy*.

Universality means just that: *the strength by which matter "couples" to the gauge field is given precisely by the conserved "charge" corresponding to the algebraic generator of the symmetry we have gauged.*

Here the gauge field is the gravitational potential and the coupling strength is simply given by the energy, as the relevant conserved quantity.

If we now turn to Yang-Mills cases, we see that for electromagnetism, the conserved quantity as derived by Noether's theorem is the algebraic generator of $U(1)$, and this is the electric charge, as befits Coulomb's law.

For flavour-$U(3)$ as a phenomenological field theory describing correctly the physics in the GeV energy region, we find that the coupling strengths to the relevant $\rho^{\mp}$, ρ°, $K^{\mp}$, K°, $\bar{K}^{\circ}$, Φ°, ω° fields are given by the corresponding proper values of the nine $U(3)$ generators when acting on the relevant matter representations. Notice that in any *internal* symmetry, the universal couplings will consist in pure numbers resulting from the algebra. Universality establishes a unique (*universal*) scale for these couplings. *In a non-gauge theory* the couplings are related within a single multiplet (say the couplings of $\pi^{\mp}$ and π° to protons and neutrons, related by an $SU(2)$ isospin global symmetry), but *there is no relation imposed by the symmetry for different multiplets. In a local gauge theory, all couplings are fixed by the algebra, whatever the multiplet.* This is precisely where universality comes in.

For colour-$SU(3)$ in QCD, the situation is more complicated because of confinement. This is because the measurement of a *charge* is generally performed by using the Coulomb force, at a distance from the source — which is impossible here.

For the (Fermi) weak force, it is given beautifully by the left-chiral currents of $SU(2)$ (which for the quarks are embedded in left-chiral-$SU(3)$ in a specific way — Cabibbo, 1963) and an additional $U(1)$, the weak hypercharge (Weinberg, 1967; Salam, 1968).

There is another aspect which played an important role in this context. Flavour-$SU(3)$ is the parity-conserving subgroup of chiral [$SU(3)$-left $\times$ $SU(3)$-right]. *The parity-conserving "charge" of the weak interaction currents coincides with a subset of the flavour-$SU(3)$ charges — a symmetry of the strong nuclear forces.* Because of this feature, the strong forces do not renormalize (i.e. do not modify the strength of) the weak vector current Fermi coupling (i.e. the weak parity preserving charge). Its value in neutron beta-decay is about the same as in muon-decay, although the first mentioned experiment involves strongly-interacting particles (neutrons and protons) whereas the second one does not — the muon is a heavy electron which does not partake in the strong nuclear interaction. The observation of this feature was instrumental in leading to the understanding of the gauge structure of the weak currents (Gershtein and Zeldovich, 1955; Feynman, Gell-Mann, 1958; Sudarshan and Marshak, 1958).

A similar effect was discovered for the parity-violating charges in the weak interactions. They also involve a subset of the chiral $SU(3)$-left $\times$ $SU(3)$-right charges, i.e. again a phenomenological symmetry of the strong interactions. These charges correspond to the part of chiral symmetry that is spontaneously broken through the Nambu-Goldstone mechanism, with the pions and kaons as Goldstone particles. As a result, one can relate the strength of the weak axial-vector charges (the parity-violating subset) to the coupling strengths of the pions and kaons (Goldberger and Treiman, 1958) which are given by chiral symmetry. In fact, it is a a certain adaptation of the universality idea that gives these values of the strengths of the couplings of the mesons $\pi^{\mp}$, π°, $K^{\mp}$, K°, $\bar{K}^{\circ}$, η° to nuclear matter, in terms of the anti-correlated part of the chiral $SU(3) \times SU(3)$ charges.

4.6 The equivalence principle

In gravity, as enunciated by Einstein, the principle of equivalence is a prescription for a transformation that does away locally (i.e. only at one selected place in each case) with the gravitational field. Geometrically, this consists in going over to a frame that is flat (that is, no apparent gravity), i.e. to the tangent manifold at that specific point.

Physically, we have to go over to an accelerated frame. Einstein himself discussed the fact that it is impossible to distinguish locally between a gravitational field and an accelerated frame. When waking up in a spaceship and feeling a strong pull towards "the floor", we do not know whether this means that the ship has arrived and is firmly anchored to a planet whose gravity we experience — or whether instead we might not just be in an accelerating phase in the trip, which glues us to that floor.

Equivalence is thus yet another symmetry between two situations, holding locally only, and specific to local gauge interactions.

In principle, such a transformation should exist for any local gauge theory. The difficulty resides in performing a local active gauge transformation that will result in the cancellation of the Yang-Mills potential. In electromagnetism, this means arranging for the appropriate phase change, for instance. In a flavour-$SU(3)$ gauge, cancelling a $\rho^{\mp}$ potential can only be an idealized cancellation since the $\rho^{\mp}$ is a massive particle and will not just vanish when its dynamical action is cancelled. This is also true with the weak interaction gauge $SU(2) \times U(1)$, mediated by the very massive (80-90 GeV) $W^{\mp}$ and Z° bosons, which we discuss in the next section.

4.7 Spontaneous symmetry breakdown in a gauge interaction

The superconductivity-inspired mechanism for a *spontaneous* breakdown of a global symmetry can be adapted to a local gauge symmetry (Higgs, 1964; Englert and Brout, 1964). *Spontaneity* implies a preservation of the conservation laws generated by the symmetry in the global symmetry case, with the vacuum and other particle states displaying the breakdown, and with related massless *Goldstone mesons*.

In the presence of a local gauge symmetry, the symmetry breakdown involves a meson multiplet, with a specific self-interaction and some dynamical features that make out of one of the components an effective false vacuum, i.e. a state with the lowest energy. This component has to be electrically neutral and should also have no other completely conserved charge, since that charge would thus communicate with the vacuum and its related symmetry would thus be spontaneously broken.

The Higgs meson multiplet — like any other matter multiplet carrying the local charge of symmetry (in the weak interaction this is $SU(2)$-left $\times$ $U(1)$) — contributes to the current and interacts with the gauge fields. As a result, some of its components (other than the *false vacuum*) undergo a reconstituting process and transmute into *third components* of the gauge field: without this mechanism, the GF is massless (like in electromagnetism) and thus has only two *polarizations* per group parameter. Now it acquires a third (longitudinal) component and becomes massive.

The original false vacuum component also acquires mass and in this simplest model of the mechanism, it should be observed as a massive spin 0 particle.

All of this has been vindicated in the weak interaction. It was possible to evaluate from the model what the masses of the gauge fields $W^{\mp}$ and Z° should be, and they were indeed observed experimentally in 1982. As of the writing of these lines, the search for the Higgs particle is on. The simplest model, however does not provide a method for the evaluation of its mass.

CHAPTER 5: SYMMETRY, ORDER, AND INFORMATION

5.1 More symmetry implies less information

Note already *the negative correlation between symmetry and information*. Symmetry represents a *lack of information*, an impossibility to specify, to provide identification, which is an important type of information.

Lack of information in large ensembles is traditionally connected with *entropy*, *disorder*. However, this statement is not precise enough. Missing information may be connected with disorder, in the sense that it becomes too difficult to specify that information because it relates, for example, to myriads of turbulent molecules. In computer language, it would involve myriads of information bits. This type of lack of knowledge is described as *subjective* because it is due to our own limitations.

But *in quantum mechanics*, on the other hand, *missing information just corresponds to its inexistence* — the physical state has not yet been generated, as long as a measurement has not been performed (a *measurement* in the sense of an irreversible interaction with a macroscopic system). At this stage, all there is just a wave-function, with a probabilistic interpretation. We know from the many experiments that have realized the EPR idea (Einstein, Podolsky, and Rosen, 1935) and applied the test provided by Bell's inequalities (Bell, 1966) that there is no physically concrete *underlying reality* other than the wave-function. This lack of knowledge is then an *objective* lack of information, information that does not yet exist.

In the case of the grey cats of the French proverb, the lack of information is due to darkness — not to inexistence — i.e. to a difficulty in the acquisition of the information, resembling the case of disorder. It is subjective.

Very recently, an advance in the study of *chaotic systems* has revealed the existence of *objective entropy in non-quantum situations*. There are problems in which an infinitesimal difference in the initial conditions will lead to totally different evolutions of the systems. These are then *unstable* initial conditions, generated in collective states by the internal interactions between the constituents. The phenomenon of turbulence in a liquid or in a gas is one such situation.

5.2 Measures of entropy

The entropy of a symmetry is the magnitude of the "Whittaker impotence" it represents. This can be given a quantitative definition by taking, for instance, the volume of the Lie group — or some quantity related to the group dimensionality. $SU(3)$ invariance is related to an 8-dimensional manifold. However, $SU(3)$ is a broken symmetry. It is broken through the c quark being about 30 times heavier than the a and b quarks. We think this is related to another force, the force responsible for the emergence of *generations* of quarks and leptons (particles resembling the electron). This force, which I named the *fifth interaction* twenty-five years ago (Ne'eman, 1964), therefore reduces the overall symmetry, leaving a subgroup $U(2)$ as the residual invariance. $U(2)$ has a 4-dimensional group manifold and is therefore a smaller symmetry and represents less entropy.

The study of entropy in relation with the need to describe complexity has produced in recent years completely different approaches to the objectivisation of entropy. The aim is to have a description that would represent, for instance, the complexity of a living cell or of an organism.

One such measure is *algorithmic complexity*, introduced by Kolmogorov and Chaitin independently. The quantity characterizing the complexity of the state is the length of the shortest computer program that can describe the state. It will represent the information content of that state, a kind of inverse of the state entropy.

A crystal can be described by a much shorter list of instructions than a living being (whose DNA is probably the relevant program). This means that the crystal embodies less information and has a higher intrinsic entropy than a living system. On the other hand, a gas with 10^{25} molecules could only be described by a program listing them all — the state and the design program are of the same magnitude. This would imply that the gas contains a very large amount of information — and little entropy — which is not what is meant by entropy.

This issue is resolved in a proposal due to Bennett. He measures order — the opposite of entropy — by the *logical depth* of the system. It represents the logical length of the program for the realization of the state, once the data is fed. To construct a living cell one would require an extremely long set of instructions. For a crystal, a limited number of steps would suffice. For a gas of molecules, the initial data would be of an enormous magnitude, but the instructions program would consist in a trivial "copy that data". This definition therefore does fit the concept of objective entropy.

We can adapt these concepts to symmetry. Instead of the dimensionality or volume of the group, we could measure the information content of the vacuum, i.e. of the multiplet containing the Nambu-Goldstone boson. One way of measuring this quantity could draw from the structure of the Young tableau for that representation of the group, which is similar to a computer program for its construction.

This does not appear interesting in finite-dimensional Lie groups, but something similar might be possible and helpful in infinite cases such as the presently fashionable group of conform transformations (transformations preserving angles) in two dimensions — a symmetry of the theory of the *quantum superstring* (the *string* for short) a "great hope" at present, as a candidate "theory of everything". The subject calls for further investigation.

CHAPTER 6: THE INTERACTION BETWEEN PHYSICS AND MATHEMATICS

6.1 Representation theory

The advances in the applications of symmetry in physics have boosted the progress in algebra and geometry. An important such step occurred when Wigner (1939) classified the unitary representations of the Poincaré group. This was essential to the description of particle states in physics, and later was instrumental for the construction of appropriate equations of motion. To construct the representations of the Poincaré group, Wigner developed a method that was later adapted by Mackey and other mathematicians to the construction of unitary representations of other non-compact groups.

Something similar occurred when physicists (Goshen and Lipkin, 1959; Dothan, Gell-Mann, and Ne'eman, 1965; Bohm and Barut, 1965) tried to generalize the idea of symmetry and describe spectra of excitations (bands) in hadron and in nuclear systematics by infinite unitary representations of simple non-compact groups such as $U(6, 6)$ or $SL(3, R)$. These resemble the spectra of the hydrogen atomic levels — which they then identified with a representation of $U(2, 2)$, or of the spinning-top which they identified as $SL(4, R)$, or also of the harmonic oscillator $U(1, 3)$. In these elementary problems in quantum mechanics, the dynamics are given and one calculates the spectrum of states as the solutions. In hadrons or nuclei, one observes the spectrum and the idea was to try and guess from that spectrum what the dynamics could be. This method became known as the SGA (Spectrum Generating Algebras).

On the mathematical side, this led to further knowledge about the structure and classification of such representations. In recent years the method has been further pursued for nuclei with some success (Arima and Iachello, 1975).

6.2 Infinite algebras

Infinite algebras entered physics in the study of the current algebra of $SU(3)$ or $SU(3) \times SU(3)$, when going over to the algebraic relations between the local currents of these groups, i.e. the charge and current densities at x or y that interact with the gauge field at x or y. A search for the representations of such algebras suffered from complications due to Lorentz invariance considerations (Dashen and Gell-Mann, 1966). The classification of these representations was nevertheless achieved (Joseph, 1967).

Meanwhile, a simplified *current algebra* was being investigated in mathematics (Kac, 1968; Moody, 1968). As a result of a complicated evolution in the dynamical theory of the strong interactions, the theory of the quantum string was evolved. It involved an algebra (Virasoro, 1970) which is isomorphic to the diffeomorphisms (i.e. the general coordinate transformations as in gravity) on the circle. Its representations were constructed by the physicists, using the methods of the SGA we mentioned. It was then shown (Marcus and Sagnotti, 1982) that should one need to introduce internal symmetries (Paton and Chan, 1969), this would have to be limited to certain types of groups (all related to rotations) which, for instance, could not include $U(n)$ and the observed symmetries.

In the evolution of the Kac-Moody algebras, what was missing was the representation theory. The method of *vertex operators* that had been developed by the physicists for the Virasoro algebra and the corresponding SGA of the string was now applied (Frenkel and Kac, 1980) to the Kac-Moody algebras.

In 1984, certain very promising features were proved to exist in the string, strengthening its case as a candidate theory of quantum gravity, to which it had been switched in the meantime. As a result, a model for the string as a "theory of everything" (i.e. a unification of gravity with the other interactions) was suggested (Gross, Harvey, Martinec, and Rohm, 1985). It utilized the technique evolved by Frenkel and Kac to overcome the limitations on the introduction of internal symmetries.

6.3 Supergroups, superalgebras, and supermanifolds

Superalgebras were first conceived in mathematics, in the study of algebraic deformations (Nijenhuis, 1955). Supergroups were also studied leisurely (Berezin and Kac, 1970). Berezin had in fact considerably advanced the calculus as applied to systems of anticommuting quantities encountered in the geometry of Grassmannian differential forms (Berezin, 1966). Through the advent of spinor particles such as the electron (the spin was identified in 1925) obeying Fermi (anticommuting) statistics, i.e. $a \cdot b = - b \cdot a$, the need for a better understanding of the algebraic foundations behind anticommuting fields became acute. Berezin clarified the structure and invented an *integration* operation for such quantities and later an appropriate modification of the determinant. The latter was needed for the treatment of unitarity in local gauge theories, using *ghost* fields, yet another important set of anticommuting quantities.

Independently, physicists introduced superalgebras to obtain additional algebraic constraints (Golfand and Likhtman, 1971). The idea was given a geometric interpretation (Volkov and Akulov, 1973). Meanwhile, in the study of the quantum string, the need for further algebraic constraints arose and it turned out that infinite superalgebras extending the Virasoro algebra could do precisely what was needed, i.e. rid the formalism of a tachyonic unphysical state (i.e. with imaginary mass) — (Neveu and Schwarz, 1971; Ramond, 1971; Aharonov, Casher, and Susskind, 1971).

The impact of these results was such that the idea of extending the Poincaré group into a supergroup was tried (Wess and Zumino, 1974). It seemed elegant and promising and was given a geometric structure (Salam and Strathdee, 1974). It was soon noticed that supersymmetry improved the renormalizability of a dynamical theory. In fact, there are now supersymmetrized gauge theories that suffer no renormalization, such as a local gauge theory with 4 spin ½ fermionic and 6 spin 0 bosonic matter fields. Yang-Mills theory were proved to be renormalizable ('t Hooft, 1971), but in this case the theory is simply finite and requires no renormalization. This was first shown to be true up to third order in the perturbative treatment (Grisaru, Rocek, and Siegel, 1980; Avdeev, Tarasov, and Vladimirov, 1980) and was later proved to all orders (Mandelstam, 1983).

In a Lie algebra, the algebraic generators obey commutation relations

$$[F, G] \cong F \cdot G - G \cdot F = H$$

i.e. the *Lie bracket* is realized by a commutator — the difference between the multiplication of two generators and the inversion in the order among them (remember that these are non-commutative groups; in Abelian groups this will yield $H = 0$). In a *Lie superalgebra*, there are two types of brackets. All generators are classified in two classes: bosonic (which is the only possibility in an ordinary Lie algebra), called *even*, and fermionic (this is the novelty) called *odd*. There are three possible brackets:

$$\begin{aligned} &[\text{even}, \text{even}] \to \text{even} \\ &[\text{even}, \text{odd}] \to \text{odd} \\ &\{\text{odd}, \text{odd}\} \to \text{even} \end{aligned}$$

only the third bracket is an anticommutator

$$\{A, B\} \cong H$$

A second reason for the interest in supergroups and the constraints they might impose on the spectrum was (and still is) the absence of dynamical constraints from the local gauge theory itself with respect to the spontaneous symmetry breaking Higgs spinless field. This brought about attempts to predict the composition of appropriate supermultiplets (Fayet, 1976).

A third physical motivation for supersymmetries arose in the attempts at further unification (*gauge unified theories*). In these theories, there arises a *hierarchy problem*. The symmetry has to be broken spontaneously twice, once at 10^{15} GeV, an extremely high energy (the highest energies presently available in accelerators is of the order of 10^3 GeV). This break separates the strong QCD from the weak + electromagnetic; then, at around 100 GeV, these two separate as we saw in section 4.2. It was soon found that such a hierarchical sequence was dynamically impossible: the lower energy would be renormalized automatically upwards, joining the upper one. It was shown that this result may be more generally true, with difficulties in understanding how the Higgs field of the $SU(2) \times U(1)$ breakdown does not acquire a very large mass.

As a result interest in supersymmetry has risen. The existence of such constraints could force the Higgs fields to have vanishingly small masses, like their fermionic partners in the supersymmetry multiplets: the latter could be required to stay massless by chiral symmetry. A mass of 100 GeV can be considered as almost zero, when compared to 10^{15} to 10^{19} GeV, and could result from some further symmetry breaking of the chiral symmetry. Experimentally, the search for the supersymmetric partners of all the known particles will soon start, when the accelerators will make it possible to produce particles with masses of the order of 500-1000 GeV.

Now let us turn to the impact on mathematics. When superalgebras started becoming useful, two main efforts were made by combined "task forces" of mathematicians and physicists: first, a precise study of the algebraic "rules of the game" (Corwin, Ne'eman, and Sternberg, 1975), then a rush to discover all possible simple Lie superalgebras, as Cartan had classified the semi-simple Lie algebras (Cartan, 1894). After an extensive, but rapid effort (Freund and Kaplansky, 1976), the race was won by V. Kac who managed to publish a complete classification (Kac, 1975).

The superalgebra extending the Poincaré algebra was gauged in a mode resembling the way in which gauging of the Poincaré group itself produces Einstein's gravity. The result has been the hypothetical theory of supergravity. It supplies a unification scheme and a certain reduction in the difficulties in the renormalization of gravity (but apparently insufficient for the removal of all difficulties).

6.4 Algebraic geometry and topology

Perhaps the greatest impact on mathematics coming from the advances in physics in recent years has been in the areas of differential and algebraic geometry and in topology.

While physicists such as C.N. Yang and R. Mills, Shaw, Utiyama and others were developing the idea of a gauge theory, mathematicians were inventing and studying precisely the same structure under the name of *fiber bundle manifolds* (Whitney, 1935; Hopf, 1935; Stiefel, 1936; Chern, 1944; Pontryagin, 1944).

A fiber bundle manifold $\Omega(M, G, F, \pi, \cdot)$ is constituted by a *base manifold* M (think of spacetime in a Yang-Mills internal symmetry gauge), a *structure group* G (the gauged Lie group), a *fiber* F (the representation of the matter fields), a *projection* π (such that when it acts on a point p in the fiber F it yields the point x in M underneath that $F(x)$), and an *action* of G upon the entire bundle which we shall not discuss here. The bundle is precisely the same thing as our gauge theory, but it is a geometric object! The gauge field with its compensation role is called in differential geometry a *connection*, etc.

Physicists discovered in the late sixties that the mathematicians "had already been there" (Lubkin, 1963; Loos, 1967; Wu and Yang, 1975). However, it was only after the renormalization of the Yang-Mills theory ('t Hooft, 1971a), including the case of spontaneous symmetry breakdown ('t Hooft, 1971b) and its adoption for QCD and for the weak force that physicists started looking for (quantum) solutions of the Yang-Mills equations. They discovered *monopoles* ('t Hooft, 1974; Polyakov, 1974), *instantons* (Polyakov, 1975; 't Hooft, 1976), *merons*, etc. which were important for the understanding of the theory's predictions in physics — but to the mathematicians they opened new fields (Atiyah, Hitchin, and Singer, 1977).

Two areas of mathematics that have been boosted by these results are the study of 3- and 4-dimensional manifolds and *index theorems*.

The 2-dimensional manifolds were classified long ago, and in fact that classification has been extensively used by physicists who work on the string between 1984-87 (Nelson, 1987), including results about moduli of Riemann surfaces, Teichmuller spaces, modular invariance, etc. This was due to the fact that a one-dimensional string maps a 2-dimensional "world sheet" in its time-evolution, and the applications to the calculations imply a summation over all possible (topologically different) two-spaces, i.e. surfaces of different genus. Indeed, the classification says that all two-spaces can be stretched and deformed to form a sphere, or a sphere with a "handle", or a sphere with two handles, etc. and that this is the full classification. The theory becomes more useful when working on complex spaces instead of real 2-surfaces.

Between 1982 and 1985, great advances were made in the understanding and the classification of 4- and 3-dimensional manifolds, a great open problem in topology (Freedman, 1982; Donaldson, 1983). Donaldson's results were very much influenced by the Yang-Mills solutions that physics had brought to light. For five dimensions and above, the situation is very simple and has been understood for some time. The 4-dimensional picture was a surprise. It revealed the existence of an uncountable infinity of different "exotic" spaces. The application of physics — quantum field theory — approaches has been pursued and is yielding very important results in the study of 3 dimensions as well (Floer, 1988; Witten, 1988)

A related problem is the *index theorem*. This is a method that probes the global structure of a manifold at its deepest. It exploits spinorial structures and relates to the Dirac equation. The index theorem (Atiyah and Singer, 1963) is directly related to these *instanton* solutions of the Yang-Mills equations and to supersymmetry, and here again, much progress has been achieved in recent years, using the insight provided by physics.

Summing up, we have seen in recent years a great mutual fertilization between physics and mathematics in the area of local gauge symmetry and topology.

REFERENCES

Adler, S.L. and Dashen, R.F. (1968) *Current Algebras*, New York: W.A. Benjamin Inc.

Aharonov, Y., Casher, A., and Susskind, L. (1971) *Physics Letters*, B35, 512.

Arima, A. and Iachello, F. (1975) *Physical Review Letters*, 35, 1065, 75.

Atiyah, M.F., Hitchin, N.J., and Singer, I.M. (1977) *Proceedings of the National Academy of Sciences of the U.S.A.*, 74, 2662.

Atiyah, M.F. and Singer, I.M. (1963) *Bulletin of the American Mathematical Society*, 69, 422.

Avdeev, L.V., Tarasov, O.V., and Vladimirov, A.A. (1980) *Physics Letters*, B96, 94.

Bardeen, J., Cooper, L.N., and Schrieffer, J.R. (1957) *Physical Review*, 126, 162.

Bell, J. (1966) *Review of Modern Physics*, 38, 447.

Berezin, F.A. (1966) *The Method of Second Quantization*, New York: Academic Press.

Berezin, F.A. and Kac, G.I. (1970) *Matematicheskii sbornik*, 82, 124.

Bohm, A. and Barut, A.O. (1965) *Physical Review*, B139, 1107.

Cabibbo, N. (1963) *Physical Review Letters*, 10, 531.

Cartan, E. (1894) *Sur la structure des groupes de transformations finis et continus*, [Thesis], Paris: Vuibert, 1933.

Chern, S.S. (1944) *Proceedings of the National Academy of Sciences of the U.S.A.*, 30, 269.

Corwin, L., Ne'eman, Y., and Sternberg, S. (1975) *Review of Modern Physics*, 47, 573.

Dashen, R.F. and Gell-Mann, M. (1966) *Physical Review Letters*, 17, 340.

Donaldson, S.K. (1983) *Journal of Differential Geometry*, 18, 279.

Dothan, Y., Gell-Mann, M., and Ne'eman, Y. (1965) *Physics Letters*, 17, 148.

Einstein, A., Podolsky, B., and Rosen, N. (1935) *Physical Review*, 47, 777.

Englert, F. and Brout, R. (1964) *Physical Review Letters*, 13, 321.

Fayet, P. (1976) *Physics Letters*, B64, 159.

Feynman, R.P. and Gell-Mann, M. (1958) *Physical Review*, 109, 193.

Floer, A. (1987) *Bulletin of the American Mathematical Society*, 16, 279.

Fock, V. (1927) *Zeitschrift für Physik*, 39, 226.

Freedman, M.H. (1982) *Journal of Differential Geometry*, 17, 357.

Frenkel, I.B. and Kac, V.G. (1980) *Inventiones Mathematicae*, 62, 23.

Freund, P.G.O. and Kaplansky, I. (1976) *Journal of Mathematical Physics*, 17, 221.

Fritzsch, H. and Gell-Mann, M. (1972) In: Jackson, J.D. and Roberts, A., eds., *Proceedings XVI International Congress of High Energy Physics*, Vol. 2, Batavia, Ill.: Fermilab, 135.

Gell-Mann, M. (1961) *The Eightfold Way*, Caltech Report CTSL 20, [Unpublished].

Gell-Mann, M. (1962) *Physical Review*, 125, 1067.

Gell-Mann, M. (1964) *Physics Letters*, 8, 214.

Gershtein, S.S. and Zeldovich, Ya.B. (1955) *Zhurnal eksperimental'noi i teoreticheskoi fiziki [JETP]*, 29, 698.

Ginzburg, V.L. and Landau, L.D. (1950) *Zhurnal eksperimental'noi i teoreticheskoi fiziki [JETP]*, 20, 1064.

Goldberg, H. and Ne'eman, Y. (1963) *Nuovo Cimento*, 27, 1.

Goldberger, M.L. and Treiman, S. (1958) *Physical Review*, 110, 1178.

Goldstone, J. (1961) *Nuovo Cimento*, 19, 155.

Golfand, Yu.A. and Likhtman, E.P. (1971) *Pis'ma v zhurnal eksperimental'noi i teoreticheskoi fiziki [JETP Letters]*, 13, 452.

Goshen, S. and Lipkin, H.J. (1959) *Annals of Physics*, 6, 301.

Grisaru, M.T., Rocek, M., and Siegel, W. (1980) *Physical Review Letters*, 45, 1063.

Gross, D.J., Harvey, J.A., Martinec, E., and Rohm, R. (1985) *Physical Review Letters*, 54, 502.

Gross, D.J. and Wilczek, F. (1973) *Physical Review Letters*, 30, 1343.

Han, M.Y. and Nambu, Y. (1965) *Physical Review*, B139, 1006.

Higgs, P.W. (1964) *Physics Letters*, 12, 132.

't Hooft, G. (1971a) *Nuclear Physics*, B33, 173.

't Hooft, G. (1971b) *Nuclear Physics*, B35, 167.

't Hooft, G. (1972) [Unpublished].

't Hooft, G. (1974) *Nuclear Physics*, B79, 276.

't Hooft, G. (1976) *Physical Review Letters*, 37, 8.

Hopf, H. (1935) *Fundamenta Mathematicae*, 25, 427.

Joseph, A. (1967) *Communications in Mathematical Physics*, 19, 106.

Kac, V.G. (1968) *Mathematics of the USSR: Izvestiya*, 2, 1271.

Kac, V.G. (1975) *Functional Analysis and its Applications*, 9, 263.

London, F. (1927) *Zeitschrift für Physik*, 42, 375.

Loos, H.G. (1967) *Journal of Mathematical Physics*, 8, 2114.

Lubkin, E. (1963) *Annals of Physics* [New York], 23, 233.

Mandelstam, S. (1983) *Nuclear Physics*, B213, 149.

Marcus, N. and Sagnotti, A. (1982) *Physics Letters*, B119, 97.

Moody, R.V. (1968) *Journal of Algebra*, 10, 211.

Nambu, Y. (1960) *Physical Review Letters*, 4, 380.

Nambu, Y. and Jona-Lasinio, G. (1961) *Physical Review*, 122, 345 and 124, 246.

Ne'eman, Y. (1961) *Nuclear Physics*, 26, 222.

Ne'eman, Y. (1964) *Physical Review*, B134, 1355.

Ne'eman, Y. (1967) *Algebraic Theory of Particle Physics*, New York: W.A. Benjamin Pub., 334 pp.

Ne'eman, Y. and Kirsh, Y. (1986) *The Particle Hunters*, Cambridge: Cambridge University Press, 272 pp.

Nelson, P. (1987) *Physics Reports*, 149, 337.

Neveu, A. and Schwarz, J.H. (1971) *Nuclear Physics*, B31, 86.

Nijenhuis, A. (1955) *Proceedings of the Royal Netherlands Academy of Sciences*, A58, 3.

Noether, E.A. (1918) *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen: Mathematisch-physikalische Klasse*, 1918, 235.

Paton, J.E. and Chan, H.M. (1969) *Nuclear Physics*, B10, 516.

Politzer, H.D. (1973) *Physical Review Letters*, 30, 1346.

Polyakov, A. (1974) *Pis'ma v zhurnal eksperimental'noi i teoreticheskoi fiziki [JETP Letters]*, 20, 430.

Polyakov, A. (1975) *Physics Letters*, B59, 82.

Pontryagin [Pontriagin], L.S. (1944) *Doklady Akademii Nauk SSSR*, 43, 91.

Ramond, P. (1971) *Physical Review*, D3, 2415.

Salam, A. (1968) In: Svartholm, N., ed., *Elementary Particle Physics: Proceedings of the 8th Nobel Symposium*, Stockholm: Almqvist and Wiksell, 367.

Salam, A. and Strathdee, J. (1974) *Nuclear Physics*, B76, 477.

Shaw, R. (1955) [Ph.D. dissertation], Cambridge University, [Unpublished].

Stiefel, E. (1936) *Commentarii Mathematici Helvetici*, 8, 3.

Sudarshan, E.C.G. and Marshak, R.E. (1958) *Physical Review*, 109, 1869.

Virasoro, M.A. (1970) *Physical Review*, D1, 2933.

Volkov, D.V. and Akulov, V.P. (1973) *Physics Letters*, B46, 109.

Weinberg, S. (1967) *Physical Review Letters*, 19, 1264.

Wess, J. and Zumino, B. (1974) *Physics Letters*, B49, 52.

Weyl, H. (1918) *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, 1918, 465.

Weyl, H. (1929) *Zeitschrift für Physik*, 56, 330.

Whitney, H. (1935) *Proceedings of the National Academy of Sciences of the U.S.A.*, 21, 462.

Wigner, E.P. (1939) *Annals of Mathematics*, 40, 149.

Witten, E. (1988) *Physics Letters*, B206, 601.

Wu, T.T. and Yang, C.N. (1975) *Physical Review*, D12, 3845.

Yang, C.N. and Mills, H. (1954) *Physical Review*, 96, 192.

Zweig, G. (1965) In: Zichichi, A., ed., *Symmetries in Elementary Particle Physics*, New York: Academic Press, 192.