

SYMMETRY AS THE LEITMOTIF AT THE FUNDAMENTAL LEVEL IN THE TWENTIETH CENTURY PHYSICS

Yuval Ne'eman

Address: Raymond and Beverly Sackler Faculty of Exact Sciences, Tel-Aviv University, Israel 69978;
Centre for Particle Physics, University of Texas, Austin, Texas 78712.

Abstract: Invariance considerations gradually entered physics in the 17th-19th centuries, but it was only in the early 20th century that (geometrical) symmetries became the dominant element. In Einstein's Special Theory of Relativity, the Poincaré global symmetry group rules over the kinematics. The General Theory of Relativity is "spanned" by two principles, Equivalence and Covariance, corresponding to local symmetries, a symmetry under active Lorentz transformations on local frames and a symmetry under passive local Diffeomorphisms. Local symmetries involve transformations, which, when performed at different locations, can be implemented at each such position to an arbitrarily different quantitative extent, thus imposing strains in the medium, known in geometry as curvature, torsion, etc. Local symmetries thus involve dynamics. The linkage between Symmetry and Conservation Laws was clarified in two theorems due to Emmy Noether (1918).

Outside of Relativity, symmetries played a minor role between the establishment of Quantum Mechanics in 1925 and our discovery of unitary symmetry in 1961. Our identification of SU(3) as the key to the hadron spectrum lead to the uncovering of yet another layer in the structure of matter, that of the quarks (1962-69). The inobservability of free quarks and the spin-statistics correlations of the quark model led to Quantum Chromodynamics as the basic Strong Interaction. Experimental studies of the Weak Interactions (1956-58) showed them to be induced by conserved charges corresponding to local gauge symmetries. Weinberg and Salam reconstructed the parent Electroweak local gauge symmetry, whose spontaneous breakdown around

.2TeV yields both electromagnetism and the Weak interactions. All of this was melted in a grand synthesis, the Standard Model (1975), a quantum theory of everything but gravity, entirely based on local gauge symmetries and covering the entire range of explored physics (<1 TeV). A parallel process of further geometrization has thereby characterized the emergence of symmetry as the dominating factor in Fundamental Physics, in the second half of the century.

1 THE BACKGROUND

Geometrical considerations were prominent in the Pythagorean doctrines - presented with the enthusiasm generally ascribed to a *cult*. The Pythagorean Philolaus was one of Democritus' teachers and may be responsible for the latter's eventual invocation of geometrical features as the sole characterization of atoms, the latter concept being perhaps inspired by his main other teacher, Leucippus. Plato laid the foundations for a view of space in which this is more than the set of relations between objects and for a rationalized geometrical view of the physical world. It did not catch with Aristotle. From Aristotle to Einstein, geometry was excluded from physics. Newton's contributions were direct and concrete, his metaphysical interests notwithstanding. His First (inherited from Galileo) and Third laws, however, could clearly be enunciated as *invariance principles*. What was missing was the *metatheory*. Huyghens' shortest path for light was another close hit, but it took the philosophical turn of mind of a Leibniz for a metatheory to be conceived. Voltaire distorted the idea and poked fun at "the best of possible worlds" - but Leibniz, whose independent invention of the calculus was published under the title "Of Maxima and Minima", was clearly laying the foundations of our metatheory - the *Action Principle*. Johann Bernoulli first managed to provide Leibniz's ideas with a concrete mould. From him to Euler, to Maupertuis (badly maligned by Voltaire for having "stolen" Leibniz's idea...), to Lagrange, to Hamilton and to Jacobi - and the "best of possible worlds" had somehow become the laziest - the one with *the least action*... The mathematical toolkit was, however, incomplete. *Group theory* still had to be invented by these two tragic teenager figures - Evariste Galois and Hendrik Abel. Gauss and his student Riemann still had to develop *differential geometry*, Darboux brought it closer to algebraization, Felix Klein and Sophus Lie launched the *Erlangen program* in 1872, and the wedding of geometry with group theory was finally consummated by the beginning of the twentieth century. Einstein's two *Theories of Relativity*, his 1905 *Special Theory* and his 1915 *General Theory* suddenly hit physics with this new instrumentation - and physics has never been the same. Emmy Noether (1918) extracted the essence of the application of the new tools in her two theorems.¹

2. EINSTEIN'S TWO SYMMETRIES – THE GLOBAL AND THE LOCAL – AS MODELS

Einstein's Special Theory – “shocking” as it was – amounted to a kinematical prescription, the invariance of the action (and later of the S-matrix) under the *Poincaré group* (or the *inhomogeneous Lorentz group*). It was first formulated ungeometrically – until Hermann Minkowski, in a 1908 address, brought out the geometrical context.² We have known ever since that in the approximation of weak gravity, we exist in a 4-dimensional spacetime with a pseudo-Euclidean metric, the *Minkowski metric* $(-1,1,1,1)$ (time is here $x^0 = ct$, c the velocity of light). The application of the Poincaré group as a symmetry of the kinematics is *global*, in the sense that the same transformation is assumed to hold everywhere, at least in its *passive* implementation, i.e., on the coordinates. In the *active* mode, the initial conditions may restrict the implementation to a single body, assuming the rest of the universe to be empty; for a *field*, however, *active* global transformations also have to be applied everywhere equally. The *action* is required to be *globally Poincaré invariant*, as a result of the geometrical properties of the spacetime manifold in which it exists. Spacetime thus impacts the material action, but the opposite does not hold for global symmetries.

Note also that although it would seem that the direct perception of a symmetry should eliminate the possibility of its not being geometrical, non-geometrical (or *internal*) symmetries may arise indirectly, through the First Noether theorem. One might detect a non-geometrical conservation law, which would then imply the existence of a symmetry. Noether's theorem might then involve a non-spacetime manifold over which the transformations by the relevant Lie group are defined – although this is not necessary, a definition of their action on the fields being all that is really needed. Also, *geometrical* symmetries are not restricted to be spacetime-geometrical, once the arena of physics is enlarged. In quantum theory, for instance, F. London³ showed in 1928 that the electron wave function should be symmetric under rotations of its complex phase – a geometrical feature of Hilbert space! T. Kaluza and O. Klein⁴ had earlier shown how electrical charge could also be derived by adding one dimension to spacetime.

Minkowski's comment is now at the foundation of physical theory, but it was also crucial to the entire evolution of physics in this century, a real turning point. It was thanks to the availability of this algorithm that Einstein could go on and bring about coherence between the new relativistic kinematics and gravity, which in its Newtonian formulation obeys Galilean kinematics. This he did in his General Theory, in which spacetime is a curved Riemannian manifold, but with the *Equivalence Principle*

requiring it to be *locally Minkowski-flat*, i.e., to provide for invariance under *local* (homogeneous) Lorentz transformations performed over a local frame (Darboux's *repères mobiles*, "tetrads" or *vierbeins*). The *Principle of Covariance* states that there is no preferred frame or coordinate system; mathematically it imposes invariance under *local diffeomorphisms* of the Riemannian R^4 . The wording used in enunciating this principle can be extended to any *passive* symmetry – namely that there is no preferred set of coordinates. Moreover, any *active*-mode symmetry can be reformulated as *the inexistence of a preferred setting for the system*, in terms of the relevant parameter. E. Whittaker has stressed this aspect by describing such symmetry principles as *postulates of impotence*.⁵

A *local* symmetry, such as that of the Lorentz group (in fact, because of Quantum Mechanics, of its universal covering group) $SL(2, C)$, implies the freedom of performing the transformations to different quantitative extents at different locations. The simplest model is a tube-shaped object. It is invariant under 2-dimensional rotations (and under a longitudinal translation, for an infinite tube only). Rotating the entire tube as one piece represents the *global* symmetry, under which the tube is indeed invariant. However, in a tube made of rubber, we might rotate (or twist) one sector only, or two sectors by different angles. The rubbery material can absorb the stresses (*curvature*). We could also at the same time stretch the tube at one or more locations, generating *torsion* stresses. Such a *local* invariance clearly induces stresses in the medium – spacetime in the case of General Relativity, Einstein's theory of Gravity. A *local symmetry therefore produces a dynamical theory, with its field-inducing source given by the Noether-conserved charge-current, as defined by the corresponding global symmetry*. Noether's Second theorem describes this relationship. Here the matter content does impinge on the medium, spacetime. Our example of a rubbery tube, and first foregoing the longitudinal translations, comes under the algebraic-geometry notion of a *fiber bundle manifold*, with the longitudinal dimension as its *base space* and the local circles (along this axis) making up the tube as its *fibers*, each fiber thus sitting over one point of the base space. H. Weyl⁶ first successfully introduced such an object in 1929, with Minkowski spacetime as the base-manifold, and the circle carrying the quantum phase of the electron as the fiber – i.e., a phase-carrying circle at each point in spacetime, with the possibility of rotating that phase by different amounts at different spacetime locations – this is Maxwell's electromagnetic theory in its quantum version, QED (quantum electrodynamics). The physics name for a fiber bundle is a *local gauge theory*.⁷ Twenty-five years after Weyl, C. N. Yang and R. Mills⁸ constructed a generalization of Weyl's gauge theory in which the local symmetry group is nonabelian, i.e., with noncommutative transformations; as an example of such a group, take rotations in 3

dimensions. Put a book on the desk before you, face-upwards. Let x be an axis parallel to the desk length, y to its breadth, z to the vertical direction in the room. Perform a 90° rotation around x , counter-clockwise as seen from the right; the book is now standing, facing you. Now rotate it by 90° clockwise (when looking from above) around z ; the book cover is facing the left. Now go back to the initial position and perform the two rotations in the opposite order. First, the 90° z -rotation (clockwise when seen from above); the book is lying parallel to the x axis, face upwards. Now the 90° counter-clockwise rotation (as seen from the right) around x ; the book is standing on its side, the front cover facing you, but the rows of writing are in the vertical direction. The results are thus entirely different when the transformations are performed in different (“commuted”) orders. The group $SO(3)$ of rotations in 3 dimensions is indeed the smallest noncommutative group amongst the Lie groups, the groups of continuous transformations; Yang and Mills indeed used it as the model for their construction.

Einstein’s theory of gravity is more complicated algebraically – somewhat like the tube when allowing longitudinal translations, as well as the rotations. We have recently discussed⁹ the ways and extents in which and to which it can be made to look like a local gauge theory. Some modifications and generalizations, possibly representing physical reality at high energies (where gravity becomes a quantum theory, in an as yet unverified form) correspond to larger locally implemented symmetry groups^{10,11} (and *supergroups*¹²⁻¹³).

3 QUANTUM MECHANICS, STRONG AND WEAK INTERACTIONS, INTERNAL SYMMETRIES

Planck’s (1900) discovery of the quantization of *action* was a beautiful and unexpected validation of the choice of this fundamental entity to represent the material and radiative content of spacetime. Euler, Maupertuis and Lagrange were vindicated in the most direct way. Action is dimensionless, it is naturally countable in terms of Planck’s quanta. Relativistic Quantum Field Theory and Feynman’s Path Integrals – even Schroedinger’s $\psi(x)$ wave-function – all fit with this choice.

An early innovation in the domain of discrete symmetries occurred in 1927, when Dirac conceived his relativistic equation for the electron. The equation had two solutions, the one indeed representing the electron – and the other a particle with the same mass and spin, but the opposite charges (electric and leptonic – see below). Such a particle – the positron – was indeed discovered in cosmic rays within several years. It became

gradually clear that aside from space (mirror) and time (reversal) reflections, nature supplies a symmetry under a similar inversion of the electric and other charges or “internal” quantum numbers, a symmetry between *particles* and their *antiparticles*.

The 1932 discovery of the neutron ushered in two new interactions, the Strong and the Weak. The sources inducing these interactions at first appeared very ungeometrical (1932-1958). Gradually, however, a new series of conserved charges corresponding to global *internal* symmetries emerged from the study of the new interactions: *isospin* I ¹⁵, *hypercharge* Y ¹⁶, *baryon number* B ¹⁷ for the *hadrons* (particles participating in the *Strong Interactions*, the nuclear “glue” holding together protons and neutrons in nuclei), *μ -lepton number*, *e -lepton number* (and much later *γ -lepton number*) for the leptons; particles which do not “feel” the Strong Interaction. The Weak Interaction also displays a “weak” *isospin* I_w current, with a 96 percent overlap with the “strong” version in hadrons, but extending to leptons too. From 1947 to 1964, particle species – mainly the *hadrons*, the strongly interacting particles – proliferated, reaching over one hundred types.

With the discovery of the electron by J. J. Thomson in 1897 and Einstein’s 1905 explanation of the *photoelectric effect* by showing that light and any other electromagnetic radiation consist of *photons*, massless particles carrying energy and momentum, the (*quantum*) particle aspect of electromagnetism was clearly realized, though much further work had to be invested before the physics (“*Quantum Electrodynamics*”, abbreviated as “QED”, the *gauge theory* mentioned above) were completely understood in 1948. Photons are emitted or absorbed by any electrically charged particle, ensuring the system’s invariance under the local abstract rotations symmetry of the electric phase suggested by Weyl. Yukawa had conjectured¹⁸ in 1935 that, in analogy to the photons, the Strong Interaction, the nuclear “glue” is propagated by a *massive* particle field (the π mesons), thus explaining the interaction’s short range. By the time the three necessary types of π had been found (1949) in cosmic rays, new and unexpected species – the four K mesons – had been added to the list, while the list of *baryons* had gained six new members (*hyperons*), aside from the protons and neutrons making up normal atomic nuclei. The new particles are produced in strong interactions but slowly decay, their lifetimes ranging from almost 10^3 secs in the case of the neutron’s “beta decay”, down to 10^{-10} seconds for some hyperons; these decays occur via the “*Weak Interactions*” the nuclear interaction identified¹⁹ by Fermi in 1934. In addition, very short-lived hadrons, decaying through the strong interaction itself, were produced in accelerators in the fifties and sixties. These particles - known as *resonances* - were generally endowed with high values of spin angular momentum.

4 THE SEARCH FOR ORDER – STRUCTURALIST APPROACHES FAIL

The search for order was launched around 1955, after the identification of *hypercharge* (or of *strangeness* S , with $Y = S + B$) by Nakano and Nishijima in Japan and by Gell-Mann in the USA. Two very different lines of thought permeated this search, a *structuralist* approach and on the opposite side *phenomenology*. The structuralists tried to guess at the inner structure of the particles, suggesting models purporting to explain the observations. The phenomenologists collected the observational information and tried to abstract this information in a mathematical formulation, assuming that structural understanding would follow. The structuralists followed several different lines. L. De Broglie, Yukawa, D. Bohm²⁰, Corben and others tried mechanical models, *rotators*, *spinning tops*, *vibrators*, etc. The idea was that *isospin* might correspond to true physical rotations - relative to the body frame, for instance. This would imply the possibility of adding up angular momentum and isospin, of transferring angular momentum between these categories – contradicting the observations. It turned out later on that such mechanistic models do fit something else altogether, namely the high values of the *spin angular momentum* of the above mentioned *resonances*, i.e., spin proper, rather than *isospin*, which is mathematically similar but contextually totally different.

Another structuralist approach, unrelated to preconceived mechanistic models and thus much less dogmatic, searched for *constituent* models. Already in 1949, Fermi and Yang²¹ had noted that from the point of view of the quantum numbers, π mesons could be “made of” nucleons (protons or neutrons) and their antiparticles. There was no pretence at a knowledge of the interaction involved in making the compound, just a suggestion of possible compositeness of the mesons and an identification of a set of constituents containing the observed quantum numbers. With the advent of strangeness or hypercharge, Sakata²² added one hyperon (Λ^0) carrying $Y = -1$, to the proton and neutron (and their antiparticles) as prospective constituents of all hadrons.

It is interesting to note that both above schools of structuralism were influenced by Marxist philosophy. In the case of Sakata and his colleague Taketani, *dialectical materialism* was often referred to, and the opposite phenomenologist approach was accused of *positivism* or of *pragmatism*. Mention of positivism was generally accompanied²³ by a reminder of the defeat of Mach and Ostwald – who doubted the “true” existence of atoms – and whose discomfiture was hailed by Lenin.²⁴ Yet another structuralist approach was launched by G. Chew²⁵ under the title of *the bootstrap*, or

also *nuclear democracy*. This was the antithesis of the constituent approach and carried Leibnizian features. Atomism - constituent models - was rejected and it was assumed that the hadrons were made of each other, with no possibility of distinguishing between constituent and composites. My favorite analogy is that of a jungle inhabited by tigers, lizards and mosquitoes. The tigers eat the lizards, the lizards eat the mosquitoes and the mosquitoes suck the tigers' blood, infect them and devour their carcasses. As a result, the tigers are "made of" lizards, the lizards are made of mosquitoes and the mosquitoes are made of tigers... In this approach, it was assumed that the particles with their symmetries emerge as the solutions to some highly non-linear equation.

Around 1967, a plausible system of equations was suggested by Horn and Schmid and other groups, using a phenomenological formulation. The solution, found by G. Veneziano, turned out to represent the dynamics of a string-like quantum system, again reproducing the systematics of the angular momentum excitations displayed by the resonances, but not the *internal* quantum numbers such as isospin, hypercharge, etc. Moreover, it was shown that the equations and the bootstrap idea do not contradict atomism – tigers, lizards and mosquitoes can all be made of a set of common constituents (such as the true atoms, in this example). Here in the String, hadrons made of quarks (see next section) fitted perfectly, as demonstrated by Harari and Rosner. But all of this was quickly dropped around 1971, when Quantum Field Theory took over again. Note, however, that the formalism of *String Theory*, born out of the bootstrap, has survived in an entirely different context. It is now considered as a candidate fundamental theory (a "TOE", *theory of everything*) capable of describing physics at Planck energies (around 10^{19} GeV) or at very short distances, at which spacetime is conjectured to be discretized (10^{-33} cm). All of this, however, is still very speculative.

5 THE SEARCH FOR ORDER: SYMMETRY ABSTRACTED FROM PHENOMENOLOGY – DISCOVERY OF $SU(3)$

The alternative approach consisted in the identification of a Lie group as the extended symmetry group of the Strong Interactions and of the hadrons. This could be accomplished by trying to fit the observed spectrum of hadron states into the (linear) representations of that group, by comparing observed intensity rules and measured couplings with those calculated from the Clebsch-Gordan coefficients of the group, relevant to those representations. Candidate Lie groups would be those having isospin and hypercharge ($SU(2) \otimes U(1)$) as subgroups.

The first such attempts explored ever-larger rotation groups. Even as an abstract group having nothing to do with physical rotations, the isospin $SU(2) = Spin(3)$ was regarded as $Spin(3) = \overline{SO}(3)$, the *double covering group* of $SO(3)$, i.e., of the group of *unimodular orthogonal transformations* (i.e., rotations) in 3 real dimensions – rather than as the *unitary unimodular group* in 2 complex dimensions $SU(2)$. Salam and Polkinghorne²⁶ first tried embedding $SO(3) \otimes U(1) \subset SO(4)$. Schwinger and Gell-Mann (as explained by Tiomno)²⁷ used $SO(7)$, putting the 7 mesons (3 states of the π isovector, 2+2 of the K and $\bar{K}$) as a 7-vector and the 8 baryons as a $Spin(7)$ spinor. This was further extended to $\overline{SO}(8)$ and $\overline{SO}(9)$ by Salam and Ward²⁸, for various reasons.

My own approach²⁹ was both more general and more methodical. Applying Dynkin's simplified graphical version³⁰ of Cartan's classification of the *semi-simple Lie algebras*³¹ and realizing that every reaction allowed by the conservation of isospin and hypercharge indeed appears to occur, I could draw the conclusion that what we are seeking is a rank $r = 2$ algebra, i.e., only two diagonal (additive) charges. The existence of a third diagonal conserved additive quantum number (within the then known particles) would result in additional selection rules, forbidding reactions otherwise allowed by I and Y conservation. Note that this does not occur with hypercharge itself just because as we in fact already conserve both isospin and electric charge Q , we are still dealing with a $r = 2$ algebra before even mentioning hypercharge. As $Q = I_3 + Y/2$, we then automatically also conserve hypercharge a priori, so that there is no clash.

There are five $r = 2$ semi-simple Lie algebras, A_2 , B_2 , C_2 , D_2 , G_2 . D_2 is the algebra of $SO(4)$, B_2 that of $SO(5)$, $C_2 = B(2)$ is the algebra of the symplectic group $Sp(4)$, i.e., orthogonal transformations in a 4-dimensional manifold with antisymmetric metric – and it turns out that $Sp(4) = Spin(5)$. Only A_2 can accommodate all 8 known baryons (N , Λ , Σ , Ξ) in one representation, assuming the spin and relative parity of the Ξ isospinor (which was not yet measured at the time) to be $J(\Xi) = 1/2^+$ and assuming an *even* relative Σ - Λ parity – which was not the favored experimental assignment in 1961. The mesons could only be assigned to a similar octet (baryon number B lay outside of A_2), the assignment thus predicting the existence of an eighth (isoscalar) meson, denoted η , soon to be discovered indeed. A variety of experimentally evaluated baryon-meson couplings fitted the group's assignments. In the exceptional group G_2 , on the other hand, the baryons would have to be assigned to a 7-dimensional representation, excluding the Λ_0 hyperon, which would have to be reassigned to a different (the singlet) representation. The then known mesons would fit in a similar 7.

Another set of prerequisites could be derived from the Weak Interactions, assuming the Weak transitions to lie within the same algebra. Only $\Delta Y/\Delta Q = 1$ transitions are observed in the hadron current in the leptonic decays of hadrons, such as $\Lambda^0 \rightarrow p^+ + e^- + \bar{\nu}_e$ or $K^+ \rightarrow \pi^0 + \bar{e}^+ + \nu_e$; G_2 contains $\Delta Y/\Delta Q = -1$ as well, whereas A_2 does not, etc.

There was one obvious conceptual “difficulty” with the choice of A_2 or (the group) $SU(3)$ in the octet version: the nucleon, considered until then as the “fundamental” hadron (especially as the proton is the only completely stable hadron) is now assigned to a “higher” representation, not to the defining one of the group. A_2 is the generating algebra of the group $SU(3)$ (unitary unimodular matrices acting over 3 complex dimensions) and thus has a 3-dimensional fundamental representation. In fact, at the advice of Yukawa, in the same year 1960-1961 a research group at Nagoya, Japan, applied $SU(3)$ to the Sakata model, with (p, n, Λ^0) in the defining representation and relegating the Σ and Ξ to other representations. The Ξ isospinor had to be assigned $J^P = 3/2$. Returning to our *octet* model, we note that in contradistinction to the Sakata version of $SU(3)$ – and also to G_2 (where the nucleons appear in a “fundamental” septet) – *the octet does not assign a fundamental role to the nucleons and they are treated very much like composite objects.*

M. Gell-Mann independently arrived – somewhat later – at conclusions similar to mine.³² Two groups, Speiser and Tarski³³ and later Behrends et al³⁴ conducted in 1961-62 methodical but inconclusive searches. Glashow and Gell-Mann³⁵ also later conducted a methodical survey, justifying the selection of $SU(3)$.

6 UNITARY (BROKEN) SYMMETRY

Before looking in detail at the $SU(3)$ formalism, we list the key physical features relating to symmetry under a Lie group: (a) Quantum numbers, i.e., a classification in multiplets, i.e., the group representations’ carrier spaces; (b) relative couplings between particles belonging to several different multiplets; (c) “intensity rules”, i.e., ratios between processes involving particles belonging to the same multiplets; (d) ratios between magnetic moments, between electromagnetic mass differences, between weak transitions, etc. – all transitions involving an interaction with a perturbative description, breaking the symmetry in a regular fashion. In the specific case of $SU(3)$ this was also true of (and the same treatment was given to) the mass differences between the isospin multiplets, e.g., $N(939)$, $\Lambda(1, 115)$, $\Sigma(1, 195)$, $\Xi(1, 320)$ (the masses are in MeV). They seemed to correspond to an effect acting in the same direction (“ λ_8 ”) as hypercharge –

but also generated a puzzle – how could an effect due to the Strong Interactions work well perturbatively? I pointed to the solution to this puzzle in 1964³⁶ by suggesting that *the mass differences indeed represent an additional, as yet unexplored, perturbative interaction and do not originate in the Strong Interactions themselves*, a solution which was later incorporated in the “Standard Model” we describe in the last section (though very little can be said about the responsible interaction, now relegated to the “Higgs sector”).

Aside from these predictions, *if the symmetry is local, we also have the prediction of the existence of a gauge multiplet (generally $J^P = 1^-$ bosons), in the adjoint representation of the group. These couple universally to the Noether-conserved currents of the symmetry, i.e., the coupling is a matrix-element of a generator of the Lie algebra.* As a matter of fact, for such a *universal* coupling, we can derive ratios for the strength of the coupling between particles in *different* representations of the gauge group – here $SU(3)$ – to the *gauge bosons*, since these couplings are given by the gauge algebra generators in the appropriate representation. Hundreds – or perhaps thousands – of predictions, of all the above types, were derived and the results were compared with the experimental results, with amazing success.

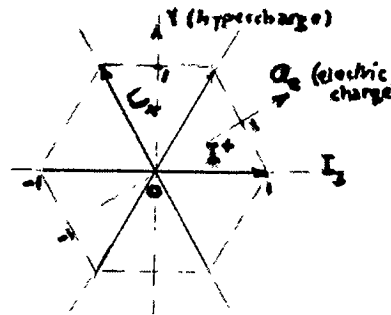


Figure 1: Root diagram of $SU(3)$

Any *simple* Lie algebra is characterized by its *root diagram*. For A_2 , the generator algebra of $SU(3)$, this is given by the six ends of a perfect hexagon, plus a double point in the center (see Figure 1). The horizontal axis represents *isospin* or I -spin; H. J. Lipkin named the two others U -spin and V -spin, following an advertisement of the times “I scream, you scream, we scream – we all scream for ice-cream.” The particle *representations’* (groupings) *weight* diagrams, giving the particle content, by displaying the particles’ values of the “additive” (diagonal) quantum numbers (for $SU(3)$) are all either triangles or hexagons.

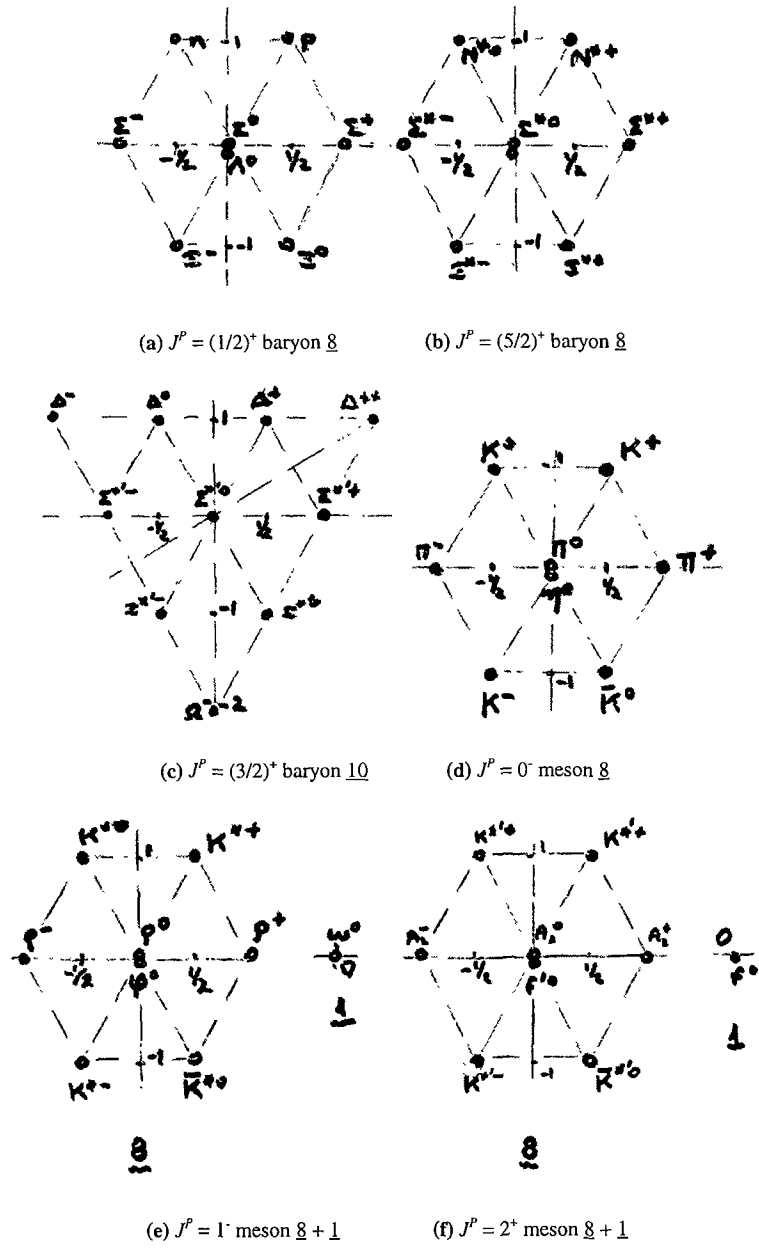


Figure 2

Figure 2 displays the weight diagrams for (a, b) the $J^P = (1/2)^+, (5/2)^+$ (hexagonal) baryon octets (the latter is a rotational-vibrational excitation of the first, an example of a “Regge recurrence”), (c) the $J^P = (3/2)^+$ (triangular) baryon decimet to which we shall return in the coming paragraphs – and (d-f) the $J^P = 0^-, 1^-$ and 2^- “octet + singlet” combination bosons.

Note that hypercharge Y is orthogonal to $\mathbf{I}$ -spin and electric charge to $\mathbf{U}$ -spin. The masses are $\mathbf{I}$ -invariant, up to electromagnetic corrections. We can immediately derive e.g., the Coleman-Glashow sum-rule for electromagnetic mass-differences. With Q orthogonal to $\mathbf{U}$ -spin, electromagnetic effects should be $\mathbf{U}$ -invariant. For the baryon $J^P = (1/2)^+$ octet, we can display the $\mathbf{I}$ -invariant masses plus the electromagnetic corrections E :

$$\begin{aligned}
 p &: m_N + E^+ \\
 n &: m_N + E^0 \\
 \Sigma^+ &: m_\Sigma + E^+ \\
 \Sigma^- &: m_\Sigma + E^- \\
 \Xi^0 &: m_\Xi + E^0 \\
 \Xi^- &: m_\Xi + E^-
 \end{aligned} \tag{1}$$

yielding the sum rule,

$$m_{\Xi^-} - m_{\Xi^0} = m_{\Sigma^-} - m_{\Sigma^+} + m_p - m_n. \tag{2}$$

Similarly, magnetic moments should be $\mathbf{U}$ -spin invariant. We get

$$\begin{aligned}
 \mu(\Sigma) &= \mu(p) \\
 \mu(\Xi^0) &= \mu(n) \\
 \mu(\Xi) &= \mu(\Sigma), \text{ etc.}
 \end{aligned} \tag{3}$$

For the different masses within a multiplet, Okubo derived the formula,

$$m = m_0 + aY + b[I(I+1) - Y^2/4] \tag{4}$$

whose specific predictions for a given multiplet may be derived from $\mathbf{U}$ -spin considerations.

This is where we return to the $J^P = (3/2)^+$ triangular multiplet (Figure 2c). The four $\Delta(I=3/2)$ states were discovered by Fermi in the fifties; the three $\Sigma(I=1)$ and the two $\Xi(I=1/2)$ states were reported at a conference at CERN (Geneva) in July 1962. At the same time, G. and S. Goldhaber reported the inexistence of similar resonances in K - N scattering (i.e., $Y = 2$). Up to this point, all the reported states could have been assigned to either one of two multiplets, the decimet of Figure 2c or a 27-state hexagonal multiplet with $I = 2$, $Y = 2$ states. The Goldhabers' negative search enabled me³⁷ to select the decimet assignment and to predict the precise properties of its missing member, the Ω^- hyperon, with $J^P = (3/2)^+$, $I = 0$, $Y = -2$, $m = 1673 \text{ MeV}$ (note that the Okubo formula simply becomes an *equal-spacing* rule, for the masses in the *triangular* decimet). M. Gell-Mann reached the same conclusions independently.³⁸ The experimental validation³⁹ of the existence and properties of the Ω^- in February of 1964 brought about the adoption of $SU(3)$ (in our baryon octet version – the Sakata model has no room for such an assignment).

7 THE QUARK MODEL

Chemistry came of age around 1868, when Mendeleev correctly identified the *order* in the patterns listing chemical elements, with their characteristic properties as abstracted from the phenomenology – *valencies* being one important example. It took, however, more than half a century and the work of Roentgen, Becquerel, J. J. Thomson, Rutherford's uncovering of the structure of the atom in 1911, Bohr's 1916 quantum version and finally Chadwick's 1932 discovery of the neutron – before the *structural* explanation of this order was understood. Biology provides us with yet another example of a classification (Carl von Linne's, in the 17th century) providing the first step in a prolonged effort, finally leading to a structural understanding (the uncovering of the genetic code in the 1950's) – with the work of Gregory Mendel representing the key intermediate stage. Progress in physics between Kepler's identification of three regularities in the motion of the planets in the solar system – and Newton's discovery of the laws of mechanics half a century later – can be considered as another such example.

The analogous sequence, from our 1960/61 classification of the hadrons, to a structural understanding of this order, appears to have gone faster. Early in 1962, with H. Goldberg, we conceived⁴⁰ a “mathematical model” which would produce the combinatorics leading to just such a pattern. (1) *The basic constituents had to be fermions with baryon charge $B = 1/3$, and (2) they would have to come in three types.* All observed baryons and their $SU(3)$ representations would thus correspond to specific

symmetry patterns of sets of three “bricks”, selected out of the “available” three types. $SU(3)$ symmetry itself is the expression of the equality between the three types with respect to the Strong Interaction, the invariance of the latter under transitions mixing or exchanging the three types. The known mesons, on the other hand, whatever their spins, are composed of a “brick-antibrick” pair. The “basic bricks” are now named *quarks*, following a suggestion by M. Gell-Mann, who about two years later provided the mathematical idea with a more physical formulation⁴¹; the name “quarks” was borrowed from a phrase in Joyce’s “Finnegan’s Wake”, “three quarks for Master Mark”. The “types” are now known as *flavors*, a term also selected by Gell-Mann, the inspiration this time coming from Baskin-Robbins, i.e., palatal rather than literary. Another presentation of the same quark-brick idea was made by G. Zweig in 1964.⁴² To display their flavors, the three quarks were denoted u , d , s (for “up”, “down”, “strange” or “sideways”, or also “singlet”). Between 1974 and 1996, three more quark flavors were discovered (the fourth was predicted) and named c , t , b (for “charm”, “top” and “bottom” or also “truth” and “beauty”). Both cosmological-astrophysical calculations and accelerator experiments appear to indicate that there are no more than six flavors altogether. The *electrical charges* of all quarks are fractional, $2/3$ for u , c , t and $-1/3$ for d , s , b , in units of e , the absolute value of the charge on the electron.

Between 1964 and 1968, the “naive” *quark model* (i.e., without a dynamical explanation) provided hundreds of experimentally verifiable predictions and passed all tests beautifully. Two simple examples relate to the ratio 3:2. F. Gürsey and L. Radicati⁴³ had suggested a methodology for non-relativistic situations (“ $SU(6)$ ”). Applied to the magnetic moments, it required the ratio between the proton’s and the neutrons’ (experimentally, both displaying major unexplained “anomalous” contributions) to have the value -3:2. The two numbers were known since the early fifties but had not been compared. The values in Bohr-magnetons are $\mu_p = 2.79$, $\mu_n = -1.96$, i.e., a ratio of -1.42. The difference fits well with the corrections due to the breaking of $SU(3)$ and the non-relativistic treatment. Another typical prediction is due to E. Levin and L. Frankfurt⁴⁴, who evaluated the ratio between the nucleon-nucleon and pion-nucleon total cross-sections, in the *high-energy* “asymptotic” limit. In this limit, the scattering of a particle and of its antiparticle on the same target should yield the same cross-sections, according to a theorem in Quantum Field Theory. In the comparison due to Levin and Frankfurt, the target is a nucleon in both cases; one scattered beam is made of three quarks, the other of quark-antiquark pairs, i.e., the equivalent of two quarks. The ratio should thus be approaching 3:2 in the asymptotic

limit. Many other predictions of that nature were published by Lipkin and Scheck⁴⁵ and others⁴⁶, all experimentally validated.

An evaluation of the “bare” masses of the quarks indicate masses $m_u = 5 \text{ MeV}$, $m_d = 9 \text{ MeV}$, $m_s = 160 \text{ MeV}$, $m_c = 1.4 \text{ GeV}$, $m_b = 5.0 \text{ GeV}$, $m_t = 170.0 \text{ GeV}$. Note that all masses are dwarfed when compared with that of the t – a mystery as yet unexplained by theory, as is the entire issue of the origin of these masses. The strong interaction appears to add about 305 MeV per quark, whatever the flavor. As a result, e.g., the proton, which is a uud compound, acquires a mass of $5 + 5 + 9 + 915 = 934 \text{ MeV}$ (the correct figure is 938 MeV), the neutron (a udd compound) should acquire a mass of $5 + 9 + 9 + 915 = 938 \text{ MeV}$ (939.6 MeV experimentally) and the Λ_0 (uds) $5 + 9 + 160 + 915 = 1089 \text{ MeV}$ for the physical $1,115 \text{ MeV}$.

The “direct” experimental proof of the existence of quarks was achieved in 1967-69 at the Stanford Linear Accelerator (“SLAC”) in deep-inelastic electron-nucleon scattering⁴⁷, somewhat in the spirit of Rutherford’s 1911 probing of the structure of the atom. At the time, Rutherford had probed a very thin metal sheet with a beam of *alpha particles* (later identified as He nuclei, with a positive electric charge $+2e$). Most projectiles went through with just a slight deflection, but in a small number of cases, the projectile was strongly reflected, almost as if it had hit very heavy and hard “rock” and was bouncing back elastically. From these results, there had then emerged a picture of an “empty” atom, with orbiting electrons – and a nucleus taking up only 1:100,000 of the radius (or 10-10 of the area) and containing all but 1:2000 of the atomic mass.

In 1957-59, Hofstadter and colleagues had probed the meson cloud surrounding nucleons, scattering on nucleon targets elastic photons emitted from recoiling electrons. Instead of the expected 0^- pion cloud, the probing (as shown by Y. Nambu) had revealed clouds of $J^P = 1^-$ mesons, surrounding the nucleon; they turned out to be coupled, in an approximation to the Yang-Mills dynamical mode for local symmetries, to $\mathbf{I}$ -spin, to hypercharge and to baryon number. These mesons were later produced directly and constitute the 1^- octet + singlet multiplet of Figure 2e, coupled a la Yang-Mills to the entire set of $U(3)$ currents, i.e., $SU(3)$ plus baryon charge. They decay into $J^P = 0^-$ mesons, which populate the outer reaches of the nucleon. In the 1967-69 version, R. Taylor and his colleagues⁴⁷ probed the “insides” of protons and neutrons by hitting them with high-momentum-transfer photons, again emitted from recoiling highly-accelerated electrons. The experiments involved large values of the momentum-transfer and *highly inelastic* scattering. The photons emitted by the recoiling electrons were found to interact with *point-like electric charges “floating” within the nucleon* and accounting for about 50 percent of the nucleons’ momentum – and with fractional values of the

charges, fitting the quark picture. The other 50 percent of the proton momentum is carried by matter which is electrically neutral.

Four puzzles still remained unresolved. The first related to the nature of the Strong Interactions *at very short distances or very high energies*, displaying what looks very much like *quasi-free* quarks, floating pointwise within the nucleon “sea”. The second related to the same subject – the nature of the Strong Interactions – but at *the low-energy or long-distance end: why are there no free quarks? Are they confined inside hadrons?* Neither do we observe two-quark or four-quark hadronic systems. This is puzzle #3: are all but $3n$ (n an integer) systems forbidden? Puzzle #4 related to *the apparent breakdown of the spin-statistics correlation in quarks*. The Ω^- ; $J^P = (3/2)^+$ hyperon (1,672 MeV), for instance, is an *sss* combination, the Δ^{++} , $J^P = (3/2)^+$ Fermi resonance (1,235 MeV) is *uuu*, etc; all also have their three quark spins aligned symmetrically (to add up to $(3/2)^+$). With their $J = 1/2$, quarks have to be fermions; how then can they form totally symmetric combinations in these hadrons?

8 QCD AND THE STANDARD MODEL

A plausible solution to all four puzzles emerged with the developments in Quantum Field Theory in 1971-73. We described in section 2 the *Yang-Mills model* of a *local symmetry*, indirectly inspired by Einstein’s General Theory though much simpler, a dynamical theory, geometrical in its nature. When it was published, the Yang-Mills “theory”⁸ was just a model, a mechanism searching for an application. In 1958-62 it was successfully applied to *I*-spin, etc., and then to the hadron’s $SU(3)$ and in this role it is an important component of the *effective* low-energy strong interaction, involving the 1⁺ mesons first encountered in the Hofstadter experiments we described in the previous section – and coupled *universally* as explained in section 6. These gauge mesons are, however, massive, though such masses break the gauge symmetry. Meanwhile, however, between 1957 and 1971, a number of theorists had worked on developing a quantized version of the Yang-Mills model, similar to QED. When this was completed in 1971, the model was restudied as to its physical features. It then turned out (1973) that these features of the exact quantized theory were just what was needed in order to understand the puzzles we listed. *These theories indeed have a coupling which weakens at shorter ranges*, where the relevant charged matter currents become effectively uncoupled. In the opposite direction, with increasing distance, the force increases like in a spring. *Quarks are thus confined*, though there is as yet no exact proof that the confinement is absolute. Another feature deserving to be mentioned: QCD is mediated

by an octet of 1^- massless (but confined too) *gluons* – explaining for instance the nature of the electrically neutral stuff found to carry 50 percent of the nucleon momentum (see section 7).

What is the charge generating the Strong Interaction, at the quark level? To explain all four puzzles, this had to be yet another $SU(3)$, which was named *color* by Gell-Mann. *Each of the six flavors of quarks appears in three colors.* This explains the *spin-statistics correlation* - the quarks enter in configurations which are *antisymmetric in the color labels*, and are thus normal fermions. It also explains *saturation at three* – the fact that configurations involving $3n$ quarks are favored energetically. *Quantum Chromodynamics (QCD)* is the name of this basic theory of the strong nuclear force. 48-49 Yukawa's original suggestion of the 0^- pion mediating the nuclear interaction – or the approximate Yang-Mills interaction mediated by the 1^- mesons – these remain true but only as “effective” higher order contributions of QCD, itself acting at the quark level. Note that $SU(3)$ flavor and $SU(3)$ color “commute”, i.e., *QCD is invariant under flavor transformations. The breaking of $SU(3)$ flavor is due to the quarks' bare masses, part of the input in QCD.*

We mentioned in sections 3 and 5 the characteristics of the Weak Interactions and their currents, such as the hadron current appearing in the neutron's *beta-decay* $n \rightarrow p + e^- + \bar{\nu}$; or the related $p + \bar{\nu} \rightarrow n + e^+$ through which Reines and Cowan proved in 1955 that neutrinos exist. This is roughly the *I-spin* transitions current ($n \leftrightarrow p$), relating to a symmetry of the Strong Interactions, which explains why this Weak coupling is unmodified by the Strong Interactions and stays equal to the one measured in muon decay $\mu^- \rightarrow \nu_\mu + e^- + \bar{\nu}_e$. It became clear in 1957 that the Weak Interactions involve currents very similar to the electromagnetic. S. Weinberg and A. Salam, using an algebraic $U(2)$ construction due to S. Glashow, raised in 1967-68 the possibility that the Weak and Electromagnetic Interactions might represent at higher energies (above 100 GeV) pieces of one single interaction.⁵⁰⁻⁵¹ The theory became precise with the progress achieved in the construction of the quantum version of the Yang-Mills model. It requires four 1^- universally coupled mesons, three of which acquire large masses when the symmetry is broken (they were indeed discovered in 1982) and the fourth – remaining massless – is the electromagnetic potential. The *electroweak theory* is thus a Yang-Mills theory with “spontaneously” broken symmetry. It turns out that such a mechanism can also be derived in a purely geometrical formulation – applying a recently developed new branch of geometry – *noncommutative geometry*.⁵²⁻⁵³ I had hit on this formulation, using “naïve” arguments⁵⁴⁻⁵⁶, before the emergence of the new mathematics; it was later shown that my formulation is that of the new geometry.⁵⁷ The

electroweak theory, proved experimentally, is thus also a dynamical and geometrical theory and forms, together with QCD, the *Standard Model*. This is a well-tested dynamical theory of the fundamental forces other than gravity, covering phenomena as explored below 500 GeV.

The entire Standard Model is thus geometrical – in line with Plato’s guess. Gravity, the remaining interaction outside it, is also geometrical – as established by Einstein – but it is as yet a *classical* theory. There are several candidate models for a quantum adaptation, but at the writing of this article, we do not yet know which is the correct version. We can only guess, at this stage, that both Quantum Gravity and beyond it, the unifying prospective theory which will encompass both that Quantum Gravity and the Standard Model (present candidates: *supergravity*, *the superstring*, “*M-theory*” or the *supermembrane*,) will also end up fulfilling Plato’s adage.

REFERENCES (NOTES AND BIBLIOGRAPHY)

- [1] Noether, E. (1918) Nach. d. Kgl. Ges. d. Wiss., *Math.-phys. Klasse* 2, 17
- [2] Minkowski, H. (1908) *Lecture at the 80th Assembly of the German Researchers and Physicians*, Cologne.
- [3] London, F. (1927) *Z. f. Phys.* 42, 375.
- [4] Kaluza, Th. (1921) *Sitz. Preuss. Akad. Wiss.*, 966, Klein, O. (1926) *Nature* 118, 516.
- [5] Whittaker, E. (1949) *From Euclid to Eddington*, Cambridge University Press.
- [6] Weyl, H. (1929) *Z. f. Phys.* 56, 330
- [7] Wu, T. T. and Yang, C. N. (1975) *Phys. Rev.*, D12, 3845.
- [8] Yang, C. N. and Mills, R. L. (1954) *Phys. Rev.* 96, 191
- [9] Ne’eman, Y. (1997) *Gauge Theories of Gravity*, Opening address at Jadwisin (Poland) symposium to be published in *Acta Phys. Polon.*, B.
- [10] Hehl, F. W., Heyde, P. v. d., Kerlick, G. D. and Nester, J. M. (1976) *Rev. Mod. Phys.* 48, 393.
- [11] Hehl, F. W., McCrea, J. D., Mielke, E. W. and Ne’eman, Y. (1995) *Phys. Rep.* 258, 1.
- [12] Freedman, D. Z., Ferrara, S. and Nieuwenhuizen, P. v. (1976) *Phys. Rev.*, D13, 3214, Deser, S. and Zumino, B. (1976) *Phys. Lett.*, B62, 335
- [13] Cremmer, E. and Julia, B. (1978) *Phys. Lett.*, B80, 48 and (1979) *Nucl. Phys.*, B159, 141.
- [14] Chadwick, J. (1932) *Proc. Roy. Soc.*, A136, 692.
- [15] Heisenberg, W. (1932) *Z. f. Phys.*, 77, 1.
- [16] Nishijima, K. (1955) *Prog. Theoret. Phys.* 13, 285; Gell-Mann, M. (1956) *Nuovo Cim. Suppl.* 2, 848.
- [17] This is generally attributed to E. P. Wigner.
- [18] Yukawa, H. (1935) *Proc. Phys. Math. Soc. (Japan)* (3) 17, 48.
- [19] Fermi, E. (1934) *Z. f. Phys.* 88, 161.
- [20] Broglie, L. et al. (1963) *Phys. Rev.* 129, 438, 451.
- [21] Fermi, E. and Yang, C. N. (1949) *Phys. Rev.* 76, 1730.
- [22] S. Sakata, *Prog. Theoret. Phys. (Kyoto)* 16 (1956) 686
- [23] Fujimoto, Y. (1970) *Foundations of S. Sakata’s Theory etc.*, *Proc. Lenin Symp.*; see also appendix by Fujimoto, Y. to M. Taketani’s letter to Nobel Committee, published in *Soryushiron Kenkyu (Study of Elementary Particles)*, Nov. 1970, p. 262.
- [24] Lenin, V. I. (1927) *Materialism and Empirio-Criticism* (vol. XIII of Coll. Works), International Pub., New York, pp. 220-230.
- [25] Chew, G. (1961) *The S-Matrix Theory of the Strong Interactions*, W. A. Benjamin pub., New York.
- [26] Salam, A. and Polkinghorne, J. C. (1955) *Nuovo Cim.* 2, 685.
- [27] Gell-Mann, M. (1957) *Phys. Rev.* 106, 1296; Schwinger, J. (1957) *Ann. Phys. (NY)* 2 (1957) 407. J. Tiomno, *Nuovo Cim.* 6 (1957) 69

- [28] Salam, A. and Ward, J.C (1961) *Nuovo Cim.* 20, 419.
- [29] Ne'eman, Y (1961) *Nucl Phys* 26, 222.
- [30] Dynkin, E B (1947) *Uspekhi Mat. Nauk.* 2, n. 4
- [31] Cartan, E. (1894) *These*, Paris.
- [32] Gell-Mann, M (1961) *Caltech rep.* CTSL-20, unp.
- [33] Speiser, D and Tarski, J (1961) IAS Princeton rep., unp
- [34] Behrends, R.E., Dreitlein, J, Fronsda, C and Lee, B.W. (1962) *Rev Mod. Phys.* 34, 1.
- [35] Glashow, S. and Gell-Mann, M. (1961) *Ann. Phys (NY)* 15, 437
- [36] Ne'eman, Y. (1964) *Phys. Rev.*, B134, 1355.
- [37] Goldhaber, G. (1985) The encounter in the bus, in *From SU(3) to Gravity*, Gotsman, E. and Tauber, G., eds., Cambridge University Press, p. 103-106; see also Cahn, R.N. and Goldhaber, G. (1988) *The Experimental Foundations of Particle Physics*, Cambridge University Press, p.116.
- [38] Gell-Mann, M. (1962) *Proc 1962 (Geneva) ICHEP*, p. 805.
- [39] Barnes, V.E. et al. (1964) *Phys. Rev. Lett.* 12, 204.
- [40] Goldberg, H. and Ne'eman, Y. (1963) *Nuovo Cim.* 27, 1.
- [41] Gell-Mann, M. (1964) *Phys. Lett.* 8, 214.
- [42] Zweig, G (1964) CERN rep TH401, unp.
- [43] Gürsey, F. and Radicati, L (1964) *Phys. Rev. Lett.* 13, 173
- [44] Levin, E M. and Frankfurt, L L (1965) *JETP Lett.* 2, 105
- [45] Lipkin, H.J. and Scheck, F. (1966) *Phys Rev Lett.* 16, 71; see also Lipkin, H J. (1973) Quarks for Pedestrians, *Phys Rep* 8, N3
- [46] See for example Cabibbo, N , Horwitz L.P and Ne'eman, Y (1966) *Phys. Lett* 22, 336.
- [47] Bloom, E.D. et al. (1969) *Phys. Rev. Lett.* 23, 930; Breidenbach, M. et al. (1969) *Phys. Rev. Lett.* 23, 933.
- [48] Fritzsche, H. and Gell-Mann, M (1972) in *Proc. XVIth ICHEP (Chicago)* 2, 135.
- [49] Weinberg, S (1973) *Phys. Rev. Lett.* 31, 494.
- [50] Weinberg, S. (1967) *Phys. Rev. Lett.* 19, 1264.
- [51] Salam, A. (1968) in *Elementary Particle Theory*, Svartholm, N., ed., Almqvist Vorlag AB, Stockholm.
- [52] Quillen, D (1985) *Topology* 24, 89
- [53] Connes, A. (1994) *Noncommutative Geometry*, New York: Academic Press, p. 661; Connes, A. and Lott, J. (1990) *Nucl. Phys.*, B18 (Proc. Suppl.), 29.
- [54] Ne'eman, Y. (1979) *Phys Lett*, B81, 196; Fairlie, D.B (1979) *Phys Lett*, B82, 97
- [55] Thierry-Mieg, J. and Ne'eman, Y. (1982) *Proc. Nat. Acad. Sci.*, USA, B79, 7068.
- [56] Ne'eman, Y and Sternberg, S. (1990) *Proc. Nat. Acad. Sci.*, USA, 87, 89.
- [57] Coquereaux, R., Haussling, R., Papadopoulos, N.A. and Scheck, F. (1992) *Int. J. Mod Phys.*, A7, 2809.