SLAC-PUB-2363
July 1979
(T/E)

QUARKS AND LEPTONS: THE GENERATION PUZZLE *

Haim Harari †

Stanford Linear Accelerator Center

Stanford University, Stanford, California 94305

Presented at the Einstein Centennial Symposium,

Jerusalem, Israel, March 14 - 23, 1979.

---

*Work supported by the Department of Energy under contract number DE-AC03-76SF00515.

†On leave of absence from the Weizmann Institute of Science, Rehovot, Israel.

1. Introduction: "Standard Wisdom"

A well-defined "standard view" of the world of quarks and leptons now exists. Much of it has been brilliantly confirmed by experiment. Some of it is yet to be confirmed, but most of us believe that this is just a matter of time. Beyond the "standard" picture, we face a long list of crucial questions, about which we know very little. The exciting physics of the next decade will probably focus on these questions. We devote this talk to a discussion of some of them.

The first generation of quarks and leptons is, undoubtedly, the best studied. We know that the left-handed (u,d) quarks and (ν_e, e^-) leptons form doublets of the electroweak $SU(2) \times U(1)$ gauge group[1] and that their right-handed counterparts are in singlets of the same group.[2] We know that the quarks come in three colors and believe that they interact with gluons, presumably according to the rules of QCD. We believe that W^+, W^- and Z are the gauge bosons of the weak interactions and that the weak and electromagnetic couplings are related by the parameter $\sin^2\theta_W \sim 0.23$.

We also know that a second and, probably, a third generation of quarks and leptons exist. All their known properties are consistent with those of the first generation, but many experimental facts are yet to be confirmed. The t-quark is still to be discovered, the b-quark and τ-neutrino are only indirectly "observed", the electroweak properties of b are not known and even the right-handed c, s, μ and τ are not fully investigated.[3] Nevertheless, it is very likely that they will all turn out to reproduce the properties of the first-generation fermions.

The "standard" description then consists of three generations of fermions (each containing two quarks and two leptons), and twelve gauge bosons: eight SU(3) gluons and four $SU(2) \times U(1)$ electroweak vector bosons. All their interactions are specified by the overall $SU(3) \times SU(2) \times U(1)$ gauge group. The free parameters of the theory include the masses of all quarks and leptons, the generalized Cabibbo angles and the coupling constants of the three gauge groups or, alternatively: α, α_s, $\sin^2\theta_W$.

So much for the "standard wisdom".

2. The Generation Pattern: Unlikely Alternatives

The presently accepted pattern of generations has two independent striking features:

(i) Within each generation, the pattern of quarks is very similar to the pattern of leptons.

(ii) Each generation is similar to the other generations.

Each of these features leads to interesting implications. The first suggests a profound connection between quarks and leptons. The second indicates that the old $e-\mu$ puzzle is now generalized into a puzzle of apparently redundant generations of both leptons and quarks which, like e and μ, differ from each other only by their masses.

It is still possible, however, that the correct pattern is different. For instance, we cannot completely exclude the possibility that each generation actually contains, say, three quarks of charges 2/3, -1/3, -1/3 (e.g.: u,d,b; c,s,h). This would break the quark-lepton similarity, unless there are two charged leptons for each neutrino.

It is also possible that future experiments will show that different generations have different structures. Higher generations may contain more fermions or they may involve right-handed $SU(2) \times U(1)$ doublets, et cetera. An even wilder possibility is the existence of "exotic" quarks and leptons. These might include doubly charged leptons and/or quarks with charges 5/3 or -4/3, and/or spin 3/2 quarks and leptons and/or color sextets.

There is no experimental evidence or theoretical need for any of the above suggestions. However, we should constantly keep an open eye for any hints in such unconventional directions. While we do not understand the pattern of identical generations, we at least have a well defined puzzle. Any deviations from the standard pattern will radically change our puzzle.

3. The Electroweak Group: Interesting Extensions of $SU(2) \times U(1)$

The experimental evidence for the validity of $SU(2) \times U(1)$ as the correct gauge theory for electroweak interactions is quite impressive. We have no reason to doubt it. However, it is entirely possible that some higher gauge group G provides a full description of the electroweak interactions, and contains $SU(2) \times U(1)$ as a subgroup. Those gauge bosons of G which lie outside $SU(2) \times U(1)$, must be heavier than W^+, W^- and Z. All the present phenomenological studies of $SU(2) \times U(1)$ could then remain essentially unchanged.

Do we have good theoretical reasons to go beyond $SU(2) \times U(1)$? There are at least three such reasons and they are related to the three obvious open questions of $SU(2) \times U(1)$:

(i) How (or why) is parity violated, leading to a very different $SU(2) \times U(1)$ classification of left-handed and right-handed fermions?

(ii) What determines the value of $\sin^2\theta_W$ which, in $SU(2) \times U(1)$, remains a free parameter?

(iii) A third possible motivation might be to include different generations in one large gauge multiplet. We will return to it in Section 8.

The question of parity is extremely interesting. There are two rather simple "orthogonal" views. One possibility is that parity is fundamentally broken at all momenta and distances. There is no energy scale in which the electroweak interactions conserve parity, and there is always a difference between the response of left-handed and right-handed fermions to the electroweak bosons. This view does not explain how or why parity is violated. It fits well with the apparent masslessness of the neutrinos (without explaining it, of course). If this approach is correct, there is no need to extend the electroweak group beyond $SU(2) \times U(1)$.

The opposite view is that at very short distances and large momenta, parity is actually conserved. The full electroweak group conserves parity in the symmetry limit and is, therefore, larger than $SU(2) \times U(1)$. Parity is spontaneously broken and its observed violation at present energies results from the fact that left-handed and right-handed fermions couple (with <u>identical</u> couplings!) to gauge bosons of <u>different</u> masses. The violation of parity is introduced via the mass spectrum of the gauge bosons, and is triggered by the same mechanism which produces the fermion and boson masses and the Cabibbo angles. The simplest group[4] which may accomplish this task is $SU(2)_L \times SU(2)_R \times U(1)$. Its gauge bosons are $W_L^{\pm}$, Z_L, $W_R^{\pm}$, Z_R, γ. If we identify $W_L^{\pm}$ and Z_L with the "usual" $W^{\pm}$ and Z, and if $W_R^{\pm}$, Z_R are substantially heavier, the entire $SU(2) \times U(1)$ phenomenology remains unchanged, except for small corrections. At the same time, we have a parity conserving theory of electroweak interactions, at energies well above the masses of $W_R^{\pm}$ and Z_R. Present data place the lower limit on these masses around 300 GeV or so. It should be interesting to improve the accuracy of the "old" β-decay and μ-decay parameters in order to increase these limits (or to discover right-handed charged currents). In such left-right symmetric theories, a massless neutrino is extremely mysterious and somewhat unlikely. However — small neutrino masses cannot be experimentally excluded.

We believe that $SU(2)_L \times SU(2)_R \times U(1)$ is an attractive possibility and that its theoretical and experimental implications should be further studied.

Our second motivation for extending the electroweak group beyond $SU(2) \times U(1)$ is the desire to calculate $\sin^2\theta_W$. Putting it more bluntly: $SU(2) \times U(1)$ is not a true unified theory of electromagnetic and weak interactions because it still has two independent coupling constants. The simple solution to this problem would be to embed $SU(2) \times U(1)$ in a larger simple Lie group which has only one coupling constant and, therefore, determines $\sin^2\theta_W$. We refer to such a theory as "simple unification" (as opposed to "grand unification" on one hand, and to $SU(2) \times U(1)$ on the other hand).

4. Simple Unification: An Attractive Idea Which Does Not Work

In order to find a "simple unification" scheme, we must seek an electroweak gauge algebra G which has the following properties:

(i) G is either a simple Lie algebra or a direct product of isomorphic Lie algebras having identical coupling constants. In both cases all couplings are defined in terms of one overall parameter.

(ii) G contains $SU(2) \times U(1)$. The gauge bosons of G which are outside $SU(2) \times U(1)$ are necessarily heavier than W^+, W^- and Z. However, there is no reason to believe that they are superheavy (say, 10^{15} GeV) and we assume that their masses are, at most, a few orders of magnitude above 100 GeV (say, less than 100 TeV). In such a case, coupling constant relations predicted by the gauge symmetry are likely to remain essentially unchanged, when tested at present energies. (This would not be the case if the additional bosons were superheavy, as they are in grand unification theories.)

(iii) Since G is "only" unifying the electromagnetic and weak interactions, G commutes with color SU(3). It, therefore, cannot relate quarks to leptons.

The idea of "simple unification" is very attractive. It would provide for a true and complete electroweak unification and would uniquely determine $\sin^2\theta_W$. This could then be the starting point for attempts to connect quarks and leptons or for schemes of unifying strong and electroweak interactions.

Unfortunately, simple unification does not work. It has been shown[3,5] that, if all quarks have charges 2/3 and -1/3, simple unification leads to $\sin^2\theta_W = 3/4$ or $3/8$, in clear disagreement with experiment. Simple unification also necessitates a pattern of quarks which is completely different from that of leptons and leads to unpleasant flavor changing neutral currents. Simple unification could conceivably be made to work if the extra gauge bosons are superheavy. However, we do not know of any reason to make such an assumption (as long as the strong interactions remain unrelated).

The failure of simple unification teaches us an extremely important lesson. It seems that a step-by-step approach may not work, while a "catch-all" solution is more successful. It is likely that complete unification of electromagnetic and weak interactions is more difficult than strong-electromagnetic-weak unification. It also appears that our ability to calculate $\sin^2\theta_W$ may depend in a peculiar way on the existence and properties of the strong interactions.

5. The Strong Group: Unlikely Alternatives to SU(3)

We have rather convincing (although indirect) evidence for the existence of three colors of quarks. We may be on the verge of obtaining evidence for the existence of gluons (more precisely — gluon jets). Quantum-Chromodynamics is far from being confirmed experimentally (in spite of many unsubstantiated claims) and is even more further away from being fully understood theoretically. However, it is a beautiful theory, essentially without competition. The features of perturbative QCD at high momenta are very attractive and are in qualitative agreement with observations. The relevance of nonperturbative effects to the question of quark and gluon confinement is less certain. In any event—color SU(3) as the gauge theory of the strong interactions appears to be an extremely good bet.

Two unlikely, but interesting, variations should be kept in mind:

(i) Color SU(3) may not be exact. It may be slightly broken, perhaps by the standard Higgs mechanism. This might provide mass to gluons, with or without affecting their alleged confinement. There is no experimental reason to suggest that color SU(3) is broken. On the other hand, the notion of a quantum number which can never be detected is perhaps somewhat chilling.

(ii) It is also possible that color SU(3) is the exact gauge subsymmetry of a larger, broken, gauge group. One candidate for the larger group is $SU(3)_L \times SU(3)_R$ with ordinary color SU(3) as the "diagonal" subgroup.[6] There is some appealing analogy between this "chiral color" and the analogous left-right symmetric electroweak group $SU(2)_L \times SU(2)_R \times U(1)$. However, the overall case for chiral color is not very convincing, in our opinion. We will return to it in Section 7.

6. Grand Unification: A Possible Quark-Lepton Connection

The analogy between quarks and leptons in each generation indicates that they must be somehow related. There are, at least, two attractive approaches to this problem:

(i) Quarks and leptons may be composite states of the same set of fundamental entities.[7]

(ii) Quarks and leptons belong to the same multiplet of a large gauge group which necessarily contains color SU(3), and therefore unifies the strong and electroweak interactions.[8]

These two possibilities are not mutually exclusive. We find the idea of composite quarks and leptons to be very attractive and we discuss it in a separate publication.[9] Here, we comment on the more popular approach of grand unification of electroweak and strong interactions.

The various motivations, the competing models and the resulting predictions and theoretical problems have all been extensively discussed[3,8] and will not be repeated here. We only wish to emphasize a few points:

(i) The choice of a grand unification scheme depends, among other things, on the choice of the full gauge groups for electroweak interactions and for strong interactions. Thus, if SU(5) is the "final word," $SU(2)_L \times SU(2)_R \times U(1)$ is excluded and parity remains violated at very short distances. Similarly, if color SU(3) is a subgroup of "chiral color", $SU(3)_L \times SU(3)_R$, most popular models are excluded. SU(5) and SO(10) are, respectively, the most natural and simple candidates corresponding to the two views of parity violation discussed in Section 3.

(ii) An attractive feature of grand unification theories is the fact that the same "superheavy" mass scale is independently calculated on the basis of two arguments. It can be estimated both from renormalization considerations of the various coupling constants and from the present experimental limit on the proton lifetime. An unattractive, unexplained, feature is the emergence of two radically different mass scales (10^2 and 10^{15} GeV) for the masses of gauge bosons. Even more unattractive and very unlikely, in our opinion, is the notion that no new physics arises between 10^2 and 10^{15} GeV, and that one can freely extrapolate over so many orders of magnitude. (The same ratio exists between the sizes of a proton and a billiard ball. Lots of things happen there.)

(iii) Most grand unification schemes do not address the pattern of generations. No known scheme can accommodate the "standard" three generations in one multiplet. This is disappointing. It may indicate, however, that the reason for generation duplication is different from the reason for a quark-lepton connection. Grand unification may be the answer to the quark-lepton similarity within a generation. It certainly does not explain the pattern of repeating generations.

(iv) It may be possible[6] to achieve grand unification at energies well below 10^{15} GeV. However, the predicted values of $\sin^2\theta_W$ seem to be too high, as we show in the next section.

(v) It is customary to assign u, d, ν_e, e to one generation and to one multiplet of SU(5) or SO(10) . However, in the same way that the "partner" of u is really $d' = d \cos\theta_c + s \sin\theta_c$, we may ask who are the partners of e^- in SU(5). In general we should define new generation-mixing angles which define the combination of u, c, t residing in the same multiplet with e^-, etc. Such angles are presumably small, but they might influence SU(5) predictions of mass relations and provide additional proton decay modes such as $\mu^+ + \pi^o$. These angles become additional parameters of the theory, on equal footing with the generalized Cabibbo angles.

7. Calculating $\sin^2\theta_W$: The Only Available Test of Grand Unification

The only numerical prediction of grand unification models which can be presently compared with experiment, and which does not depend on detailed assumptions of the Higgs structure, is the calculation of $\sin^2\theta_W$. Any model which follows the "standard wisdom" (see Section 1) without adding new quarks and leptons, gives $\sin^2\theta_W = 0.375$ at the grand unification mass. This prediction suffers substantial renormalization when we try to apply it to presently available energies, which are thirteen orders of magnitude away. The calculation of these renormalization effects is, by now, standard.[10] However, the final result depends on the "relatively low energy" subgroup of the specific grand unification scheme. The phrase "relatively low energy" refers to energies which may be a few orders of magnitude above 10^2 GeV, but are far below 10^{15} GeV. Consequently, the renormalization of $\sin^2\theta_W$ is essentially accomplished between the grand unification mass and the "relatively low energy" of the non-unified subgroup.

We may wish to consider the electroweak subgroups of $SU(2) \times U(1)$ or $SU(2)_L \times SU(2)_R \times U(1)$ and the strong subgroups of $SU(3)$ or $SU(3)_L \times SU(3)_R$. There are four combinations, leading to four expressions for $\sin^2\theta_W$:

(A) $G_{EW} \equiv SU(2) \times U(1)$; $G_S \equiv SU(3)$. This is the case for SU(5), but also for larger groups provided they break down to SU(5). The obtained expression is:

$$\sin^2\theta_W = \frac{1}{6} + \frac{5}{9}\frac{\alpha}{\alpha_s}$$

for $\alpha/\alpha_s \sim 0.05$ we get $\sin^2\theta_W \sim 0.195$.

(B) $G_{EW} \equiv SU(2)_L \times SU(2)_R \times U(1)$; $G_S \equiv SU(3)$. This is the case for SO(10), provided that $W_R^{\pm}$ and Z_R have masses of order, say, TeV rather than 10^{15} GeV. In this case:

$$\sin^2\theta_W = \frac{1}{4} + \frac{1}{3}\frac{\alpha}{\alpha_s} \sim 0.27$$

Note that an SO(10) scheme with superheavy W_R and Z_R gives the result in (A) above.

(C) $G_{EW} \equiv SU(2) \times U(1)$; $G_S \equiv SU(3)_L \times SU(3)_R$. This corresponds to an extremely unlikely situation in which we have chiral color but no left-right symmetry in the electroweak interactions. We find:

$$\sin^2\theta_W = \frac{2}{7} + \frac{10}{21}\frac{\alpha}{\alpha_s} \sim 0.31$$

(D) $G_{EW} \equiv SU(2)_L \times SU(2)_R \times U(1)$; $G_S \equiv SU(3)_L \times SU(3)_R$. This is the case for the $[SU(4)]^4$ scheme.[6] We obtain:

$$\sin^2\theta_W = \frac{1}{3} + \frac{2}{9}\frac{\alpha}{\alpha_s} \sim 0.34$$

It is clear that with the presently accepted value of $\sin^2\theta_W \sim 0.23$, cases (C) and (D) are excluded. It is the failure of this prediction which makes chiral color and early grand unification unattractive, in our opinion. Cases (A) and (B) are both acceptable.

We must repeat, however, our general reluctance to rely heavily on calculations which are based on extrapolations covering thirteen orders of magnitude.

8. What Identifies a Generation?

We do not know the reason for the existence of three similar generations of quarks and leptons. The fermions in each generation respond in an identical way to the gauge bosons of $SU(3) \times SU(2) \times U(1)$. They differ from each other by their masses and (not independently) by their couplings to the Higgs bosons. What is the secret behind the existence of generations? What defines them? Is there a quantum number which labels the generations?

One possibility is that each generation is, in some sense, an excited state of the first generation. If quarks and leptons are composite, the first generation fermions may represent the ground state of some composite system while the next generations represent higher excitations. However, these are not spin or angular momentum excitations and they cannot be radial excitations because of the relatively small mass differences between generations, as compared with the necessary large mass scale corresponding to the small dimensions of quarks and leptons. The excitations must therefore be of something else, and we do not know anything about it.

Another possibility is to suggest that there is a discrete "phase" symmetry or a $U(1)$ symmetry which act differently on different generations. This is a completely arbitrary hypothesis which explains nothing and is not motivated by any theoretical idea. However, such an assumption, together with simple constraints on the Higgs particles, leads to interesting relations between quark masses and Cabibbo angles. We discuss those in Section 10.

One may imagine that there is a "horizontal" gauge symmetry[11] among the generations. The overall gauge symmetry would then be $SU(3) \times SU(2) \times U(1) \times H$, where H acts on the generations and its quantum numbers label the generations. Such a scheme leads naturally to a duplication of generations. The number of similar generations is the dimensionality of the multiplet of H.

All horizontal models must yield flavor changing neutral currents. The gauge bosons of H are, of course, neutral and they do change flavor. Hence, they must be heavy. A particularly interesting experimental quantity related to such masses is the width for $K_L^o \to \mu^- e^+$. If we assign quantum numbers a_1, a_2, a_3 to the three generations, respectively, the simplest process which conserves this quantum number and which involves flavor changing neutral currents is $K_L^o \to \mu^- e^+$. The present upper limit on the rate yields a lower limit of 30 TeV for the mass of the gauge boson of H which connects the first generation to the second generation.

An interesting problem in horizontal gauge symmetries relates to the hierarchy of generations. Can every generation transform to every other generation by a gauge boson in H, or is there a hierarchy (e.g., only I $\leftrightarrow$ II and II $\leftrightarrow$ III transitions are induced to lowest order)? The simplest examples of these two options would be $H \equiv SU(3)$ and $H \equiv SU(2)$, respectively. The second possibility is more attractive, in our opinion, because of the apparent smaller Cabbibo mixing of "distant" generations (I and III). The SU(2)-group has another advantage: it has no anomalies. However, no completely satisfactory horizontal model has been proposed, so far.

An even more ambitious approach would be to embed SU(3) × SU(2) × U(1) × H (or G × H, where G is a grand unification group) in an even larger group, such that sll fermions belong to one multiplet. This seems to be impossible, if all generations have identical structure. However, there may be some clever ways around this difficulty.

A last tool which might prove useful is the permutation symmetry among generations, which is automatically contained in the Lagrangian of the full QCD + electroweak theory, except for its Higgs sector. This can shed no light on the generation pattern, but may be useful in discussing the connection between quark masses and Cabibbo angles.

All in all, the generations puzzle is well defined but no solution is in sight.

9. Quark Masses and Cabibbo Angles: The Framework

The standard electroweak gauge model envisages two logical stages of development: In the symmetry limit all quarks and leptons are massless. There is no difference between u and c, e and μ, d and s. Cabibbo angles are meaningless. All generations are equivalent.

The complete symmetry is then spontaneously broken, presumably via the Higgs mechanism. Quark and lepton mass matrices appear. If we know all the properties of all Higgs particles (their number, their representations, their vacuum expectation values, their couplings) we obtain mass terms of the form:

$$\left(\bar{u}_0 \; \bar{c}_0 \; \bar{t}_0 \right)_L \; M_U^0 \begin{pmatrix} u_0 \\ c_0 \\ t_0 \end{pmatrix}_R + \text{h.c.}$$

and similar expressions for d,s,b et cetera. The matrix M_U^o is the mass matrix in an arbitrarily chosen basis u_0, c_0, t_0. It need not be hermitian, and it can be diagonalized by a biunitary transformation, yielding the "physical" quark masses:

$$M_U = L_U^{-1} M_U^o R_U$$

$$M_D = L_D^{-1} M_D^o R_D$$

M_U, M_D are diagonal matrices with eigenvalues m_u, m_c, m_t and m_d, m_s, m_b respectively. L_U, R_U, L_D, R_D are unitary matrices. The standard generalized Cabibbo angles are contained in the matrix:

$$C = L_U^{-1} L_D$$

A complete knowledge of the mass matrices M_U^o, M_D^o determines all quark masses and all Cabibbo angles (including the CP-violating phase[12]). A complete understanding of the Higgs sector of the theory (or of whatever is the responsible mechanism for generating the masses) is necessary for a complete knowledge of the mass matrices.

In the absence of a convincing theoretical description of the physics behind the mass matrices, we are reduced to simple "games" with mass and angle parameters. If the correct number of generations is three, we have six quark masses and four angle parameters.[3] Hence, if the mass matrices M_U^o and M_D^o can be expressed in terms of less than ten parameters, relationships among masses and angles must follow.

Note that if we perform the <u>same</u> unitary transformation on M_U^o <u>and</u> M_D^o, no physical parameters change. This would only amount to a redefinition of our original arbitrary quark basis. Consequently, the number of physically meangingful parameters in M_U^o and M_D^o is smaller than would originally appear.

But why should we believe that there are mass-angle relations? A complete theory should probably enable us to calculate all quark masses as well as all angles. However, even if all quark masses are accepted as god-given parameters, one may argue that the angles should be expressed in terms of the masses. The argument runs as follows: certain low-energy quantities such as the $K_S^o - K_L^o$ mass difference and certain other weak amplitudes, are increasing functions of the masses of their intermediate quark lines (e.g., ΔM_{K^o} has a term proportional to m_t^2, et cetera). It is extremely unlikely that such low-energy quantities would dramatically change if the mass of the heaviest quark is changed. There is only one way of avoiding this and it is physically very attractive. If the squared Cabibbo-like angle connecting a heavy quark of mass m_Q to the lightest quarks is inversely proportional to m_Q, the contributions of m_Q to, say, ΔM_{K^o}, would always remain small, regardless of the value of m_Q. While we cannot express this argument in a general and rigorous way, we believe that it is essentially correct. It leads to two interesting conclusions: the elements of the generalized Cabibbo matrix must depend on the quark masses, and the off-diagonal matrix elements should be small (actually — if there are many generations, elements of the Cabibbo matrix which are further away from the main diagonal must be smaller).

These considerations lead us to suspect that relations among quark masses and Cabibbo angles may be derived by making relatively naive assumptions, even without a profound understanding of the generation structure. Many such attempts have been published.[13,14] We now discuss an interesting exercise of this nature.

10. An Interesting Quark Mass-Matrix

An amusing exercise may teach us several interesting lessons concering the quark mass matrices. Let us assume that the electroweak group is $SU(2)_L \times SU(2)_R \times U(1)$ and that the quark mass matrices M_U^o, M_D^o are real. (The latter assumption is made only for the sake of simplicity. We shall relax it later.) It is clear that the full gauge-invariant Langrangian (excluding the Higgs sector) is invariant under permutations among the different generations. In the case of n generations, we have a discrete $S_{nL} \times S_{nR}$ symmetry. We now allow a completely arbitrary "phase symmetry", which may be discrete or continuous, such that each generation of quarks transforms into itself, times a phase factor.[13]

$$q_i \rightarrow e^{i\eta_i} q_i$$

where i is the generation number and η_i is arbitrary. Such an arbitrary symmetry may or may not distinguish between some or all of the generations. Each Higgs field presumably has well defined properties under our abritrary "phase symmetry":

$$\phi_j \rightarrow e^{i\chi_j} \phi_j$$

where, again, χ_j is arbitrary. Yukawa couplings will, or course, be allowed, only if they are invariant under the "phase symmetry". If two or more generations remain indistinguishable by their η_i, we assume that their Yukawa couplings possess the residual permutation symmetry among them. Finally, we assume that the total number of Higgs multiplets which couple to quarks is, at most, two.

This set of assumptions is, of course, quite elaborate and arbitrary. It represents however, a "phenomenological" approach to the question of identifying the generations. It is quite general in the sense that many published models[13] are specific cases of our exercise.

Based on the above assumptions, we may now try to construct all possible mass matrices. It is clear that in each case we either have vanishing matrix elements (because of the "phase symmetry") or we have relations among matrix elements (because of the permutation symmetry). A careful study of all possible cases shows that there is a surprisingly small number of solutions. If we ignore "trivial" solutions (i.e., those in which at least two quark masses or at least one angle or the trace of the mass matrix vanish) we can prove[15] that there is an essentially unique form of the mass matrix.

In the case of two generations, M_U^o and M_D^o must have the forms:[13]

$$\begin{pmatrix} 0 & x_U \\ x_U & y_U \end{pmatrix} ; \begin{pmatrix} 0 & x_D \\ x_D & y_D \end{pmatrix}$$

While in the case of three generations there are two solutions. The first solution is:[13]

$$M_U^o = \begin{pmatrix} 0 & A_U & 0 \\ A_U & 0 & B_U \\ 0 & B_U & C_U \end{pmatrix} ; \quad M_D^o = \begin{pmatrix} 0 & A_D & 0 \\ A_D & 0 & B_D \\ 0 & B_D & C_D \end{pmatrix}$$

with:

$$\frac{A_U}{C_U} = \frac{A_D}{C_D}$$

The second solution is:[15]

$$M_U^o = \begin{pmatrix} W_U & X_U & 0 \\ X_U & 0 & Y_U \\ 0 & Y_U & Z_U \end{pmatrix} \quad M_D^o = \begin{pmatrix} W_D & X_D & 0 \\ X_D & 0 & Y_D \\ 0 & Y_D & Z_D \end{pmatrix}$$

with:

$$\frac{W_U}{Y_U} = \frac{W_D}{Y_D} \; ; \quad \frac{X_U}{Z_U} = \frac{X_D}{Z_D} \; .$$

If we relax the assumption of real mass matrices, the only added complications are some arbitrary phases in the mass matrices. In the case of two generations we obtain one prediction:[13]

$$\theta_c = \left| \text{arc tan} \sqrt{\frac{d}{s}} + e^{i\delta} \, \text{arc tan} \sqrt{\frac{u}{c}} \right|$$

where δ is an arbitrary phase and the quark labels denote their masses. Using the standard "current-quark" masses, this yields:

$$9^o \leq \theta_c \leq 16^o$$

In good agreement with experiment ($\theta_c = 13^o$). In the case of three generations we get from the first solution:[13]

$$\frac{uct}{(u - c + t)^3} = \frac{dsb}{(d - s + b)^3}$$

$$\theta_1 \equiv \theta_c \sim \left| \sqrt{\frac{d}{s}} + e^{i\delta} \sqrt{\frac{u}{c}} \right|$$

$$\theta_2 \sim \frac{1}{\theta_1}\sqrt{\frac{d}{s}}\left|\sqrt{\frac{c}{t}} + e^{i\zeta}\sqrt{\frac{s}{b}}\right|$$

$$\theta_3 \sim \frac{1}{\theta_1}\sqrt{\frac{u}{c}}\left|\sqrt{\frac{c}{t}} + e^{i\zeta}\sqrt{\frac{s}{b}}\right|$$

where δ, ζ are arbitrary phases and we assume $u \ll c \ll t$, $d \ll s \ll b$. These expressions give:

$$m_t \sim 13 \text{ GeV}, \quad 8^o < \theta_2 < 28^o, \quad 2^o < \theta_3 < 8^o \quad .$$

All of these predictions are consistent with the present bounds on the relevant parameters.[3] The second solution in the case of three generations predicts that in the limit of m_b, $m_t \to \infty$, the Cabibbo angle $\theta_1 = \theta_c$ vanishes while for m_u, $m_d \to 0$, the mixing between the second and third generations vanishes. We consider this an extremely unattractive feature. In contrast, the first solution gives a value of θ_c which does not depend at all on m_t, m_b and the mixing between the two higher generations is unaffected if m_u, $m_d \to 0$. We therefore discard the second solution and suggest that both for two generations and for three generations there is a unique solution. The solution has several attractive features:

(i) Each angle is inversely related to the mass of the heavy quark which it mixes.

(ii) The three-generations solution joins smoothly with the two-generations solution, both for m_b, $m_t \to \infty$ and for m_u, $m_d \to 0$.

(iii) The form of the mass matrix can easily be generalized to the case of an arbitrary number of generations, while keeping the above two features intact.

While the assumptions which led us to the derivation of these mass matrices are arbitrary and unsatisfactory, it is entirely possible that the matrices themselves are approximately correct. In fact, many authors,[13] starting from many different (and always arbitrary) sets of assumptions, have "derived" the same forms of matrices. Our more general derivation explains why such different starting points always lead to the same conclusions. However, we believe that, so far, neither we nor anyone else has shed any light on the question of identifying the physical differences among the generations.

We may have some correct relations between quark masses and Cabibbo angles, but we are far from understanding the generation structure.

11. Some Open Questions and Some Prejudices

We conclude by listing some of the central open questions of the world of quarks and leptons.

Are quarks and leptons related to each other?

Are the higher generations some kind of excitations of the first generation?

Is parity conserved at very short distances?

Are quarks and leptons composite?

Are there relationships among quark masses and Cabibbo angles?

We suspect that the answers to all of these questions are in the affirmative, but we are far from fully understanding any of them.

Other open questions involve the number of generations, the possible existence of "exotic" quarks and leptons, the absolute conservation of quantum numbers such as baryon number, lepton number and color, the calculation of $\sin^2\theta_W$ and, last but not least, the confinement of quarks and gluons. We have a full agenda for the next few years (or decades!).

References

1. S. Weinberg, Phys. Rev. Lett. 19, 1264 (1967); A. Salam, Proceedings of the 8th Nobel Symposium, Stockholm, 1968.
2. See, e.g., L. F. Abbott and R. M. Barnett, Phys. Rev. D18, 3214 (1978).
3. For a review see, e.g., H. Harari, Physics Reports 42C, 235 (1978).
4. See, e.g., S. Weinberg, Phys. Rev. Lett. 29, 388 (1972); J. C. Pati and A. Salam, Phys. Rev. D10, 275 (1974); R. N. Mohapatra and J. C. Pati, Phys. Rev. D11, 566 (1975).
5. S. Okubo, Hadronic Journal 1, 77 (1978).
6. V. Elias, J. C. Pati and A. Salam, Phys. Rev. Lett. 40, 920 (1978)
7. See, e.g., J. C. Pati and A. Salam, Phys. Rev. D10, 275 (1974).
8. J. C. Pati and A. Salam, Phys. Rev. D8, 1240 (1973); H. Georgi and S. L. Glashow, Phys. Rev. Lett. 32, 438 (1974).
9. H. Harari, "A Schematic model of Quarks and Leptons," SLAC-PUB-2310, Phys. Letters, in print.
10. H. Georgi, H. R. Quinn and S. Weinberg, Phys. Rev. Lett. 33, 451 (1974).
11. See, e.g., F. Wilczek and A. Zee, Phys. Rev. Lett. 42, 421 (1979).
12. M. Kobayashi and K. Maskawa, Progress of Theoretical Physics 49. 652 (1973).
13. See, e.g., S. Weinberg, Transactions of the New York Academy of Sciences, Volume 38 (1977); H. Fritzsch, Phys. Letters 70B, 436 (1977); F. Wilczek and A. Zee, Phys. Letters 70B, 418 (1977); H. Fritzsch, Phys. Letters 73B, 317 (1978); T. Kitazoe and K. Tanaka, Phys. Rev. D18, 3476 (1978); M. A. de Crombrugghe, Phys. Letters 80B, 365 (1979).

14. A. de Rujula, H. Georgi and S. L. Glashow, Annals of Physics 109, 258 (1977); T. Hagiwara, T. Kitazoe, G. B. Mainland and K. Tanaka, Phys. Letters 76B, 602 (1978); H. Harari, H. Haut and J. Weyers, Phys. Letters 78B, 459 (1978); G. Segre, H. A. Weldon and J. Weyers, Phys. Letters 83B, 351 (1979); E. Derman, Phys. Letters 78B, 497 (1978); S. Pakvasa and H. Sugawara, Phys. Letters 73B, 61 (1978); A. Ebrahim, Phys. Letters 76B, 605 (1978); 72B, 457 (1978); H. Georgi and D. V. Nanopoulos, Harvard Preprint (1979); R. Barbieri, R. Gatto and P. Strocchi, Phys. Letters 74B, 344 (1978).

15. H. Harari, to be published.