Attempts at a Unified Theory of Elementary Particles

Hideki Yukawa

Research Institute for Fundamental Physics, Kyoto University, Kyoto, Japan

MODERN atomic theory has been trying to draw as complete a picture of the material world as possible in terms of as few elementary constituents as possible. It seemed that we came closer than ever before to the goal in 1932, when the neutron was discovered. The electrons, protons, and neutrons turned out to be the only constituents of ordinary substances, whereas the photons were associated with the electromagnetic field. The positron was discovered in the same year, but this was welcome as the confirmation of the already successful theory of electrons by Dirac. However, it was clear that the picture was not yet complete. There were two outstanding problems: the beta decay and the nuclear forces. The success of Fermi's theory of beta decay led us to accept the existence of the neutrino, which has been postulated by Pauli. A relativistic field theory of nuclear forces led us further to another new elementary particle. The duality of field and particle seemed to presuppose the existence of the mesons which were to be associated with the nuclear-force field. One type of meson, the mu-meson, was discovered by Anderson and Neddermeyer in 1937 but later turned out to have little, if anything, to do with nuclear forces. Instead, the pi-meson, which was discovered by Powell in 1947, is the one that is responsible for a part, at least, of nuclear forces.

This appeared to be a little too complicated to be accepted as something final, but this was merely the beginning of further complications. Since 1947, unstable particles have been discovered in cosmic rays, in great variety, one after another. Some of them were created artificially by high-energy accelerators. It seems that more and more new particles are discovered as we go further and further in search of the high-energy region. It seems that we are in an open world in the sense that a small number of elementary particles that have been familiar to us are not likely to be the sole elementary constituents of our world but are more likely to be the more stable members of a large family of elementary particles. Of course, there is still room for the argument that most of the newly discovered unstable particles are not elementary but are compound systems that consist of two or more elementary particles in the true sense. However, even if we take such a conservative view of the present situation in the theory of elementary particles, we cannot help asking ourselves the question: What is an elementary particle?

Reexamination of the Concept of Elementary Particle

At first sight, there is no difficulty in defining an elementary particle in mathematical terms. In relativistic quantum mechanics, which was established by 1930 chiefly by Dirac, Heisenberg, and Pauli, the duality of wave and corpuscle is best represented by the concept of quantized field. It is the totality of infinitely many operators $\psi_\alpha(x_\mu)$, where x_μ is a set of space-time parameters and α is an index discriminating the components of a quantity such as a vector or a spinor that transforms linearly under Lorentz transformations. Let us call it a *local* field in order to distinguish it from a *nonlocal* field, which is discussed later. Now, an elementary particle could be defined as one that is associated with an irreducible local field. A field is said to be irreducible if it can no longer be decomposed into parts, each of which transforms linearly by itself under Lorentz transformations. In this way, the *spin* of the elementary particle is defined: For instance, the scalar or pseudoscalar field with only one component is associated with the particles with spin zero, whereas the spinor field is associated with those with spin ½. The commutation relationships between the quantized field quantities determine the statistics of the corresponding assembly of particles.

One of the most attractive features of quantum theory of fields was that it enabled us to deduce the well-known relationship between spin and statistics: The particles with zero or integer spin obey Bose-Einstein statistics, while those with half-integer spin obey Fermi-Dirac statistics. We take it for granted, furthermore, that each type of elementary particle has its unique mass m. The difficulty of the present field theories arises in this connection. Suppose that the free field satisfies the second-order wave equation

$$\left(\frac{\partial^2}{\partial x_\mu \partial x_\mu} - \kappa^2\right)\psi_\alpha(x_\mu) = 0 \qquad (1)$$

as usual, where κ is a constant with the dimension of reciprocal length. Then, of course, the mass of the associated particles is uniquely defined by $m = \kappa\hbar/c$ as long as the particles are completely free. However, in the present field theories, one can find no *a priori* reason for choosing one value or another for the constant κ or m. Therefore, what one does is to equate m with the observed mass of the particle in question. However, this again is objectionable because the particle in question is observable for the very reason that it is not free but interacts with other particles.

Thus, the problem of mass of an elementary particle cannot be separated from the problem of interaction between quantized fields. In the usual local field theory, we assume a *local interaction* between local fields. For instance, the effect of another field $\varphi_\beta(x_\mu)$ on the field $\psi_\alpha(x_\mu)$ could be described by adding certain terms to the left-hand side of Eq. 1, which are functions of φ_β and ψ_α at the same space-time point x_μ. If we introduce such an interaction, the mass of the particle that is associated with the field $\psi_\alpha(x_\mu)$ is altered by an amount that is c^{-2} times the self-energy. Unfortunately, the self-energies of particles turned out to be infinite or, at least, indefinite in the known simple cases of local fields with local interactions. This difficulty was known already in 1930, when quantum electrodynamics was established by Heisenberg and Pauli. As a matter of fact, at least a part of this pathological character of quantum theory of fields was inherited from its predecessor, classical electrodynamics. One must admit that the precise definition of the mass of an elementary particle is impossible, unless one is able to get rid of the infinite self-energy somehow.

Mixed Field Theories

The so-called "mixed field theory," which was proposed by Pais and Sakata, is of great interest in this connection. Let us take the familiar case of the electron interacting with the electromagnetic field. The self-energy of the electron owing to the electromagnetic field produced by the electron itself becomes infinite. However, if we assume further that the electron interacts at the same time with another field of appropriate kind in an appropriate manner, we may hope that the self-energy due to the latter interaction just counterbalances the electromagnetic self-energy of the electron so that the resulting self-energy becomes finite. This is actually the case, if we choose as the second field a scalar field, with which neutral spin-zero particles with the rest mass of the order of meson masses are associated, and which interacts with the electron as strongly as the electromagnetic field.

Moreover, if we extend the same idea to the case of the proton, we obtain the correct sign and the correct order of magnitude for the difference of the masses of the proton and the neutron. This seemed to give rise to a new hope of constructing a consistent field theory that would be free of the pathological divergence difficulties, by assuming the coexistence of a number of fields, known and unknown, in such a way that the self-energies of all the particles that were associated with these fields would become finite on account of mutual compensation. Such an attempt was successful to some extent, but there is little hope for arriving at the complete removal of all divergences as long as we hold to the local field theories with local interactions. Namely, the divergence that is related to the so-called "vacuum polarization" in quantum electrodynamics cannot be removed by the assumption of coexistence of various charged particles with different spins. In spite of this, however, the idea of mutual compensation is significant in indicating that the coexistence of various fields and particles is not accidental; rather, one may be able to find cogent reasons for it.

In passing I want to mention that a recent development in quantum electrodynamics originated by Tomonaga, Schwinger, and Feynman was remarkable in that all experimental results so far known were reproduced unambiguously and with great accuracy, but this was possible only after the theoretically infinite masses and electric charge had been replaced by the observed finite masses and charge. Complete justification for this *renormalization* cannot be found in the theoretical framework itself.

Local Fields with Nonlocal Interaction

In connection with the procedure of renormalization, the various types of local interaction between local fields can be divided into two classes. The first class includes all interactions that are renormalizable. The familiar interaction between the charged particle with spin ½ and the electromagnetic field is said to be renormalizable because the renormalization of the masses of the charged particle and photon and of the electric charge is sufficient to derive finite results for all other observable quantities. The scalar or pseudoscalar interaction between the scalar or pseudoscalar meson field and the nucleon, which is familiar in the meson theory of nuclear forces, is also renormalizable. There are a few other interactions that belong to the first class. However, most of the other interactions, such as those between electric and magnetic dipoles and the electromagnetic field or the pseudovector interaction between the pseudoscalar meson field and the nucleon, belong to the second class, because the divergences appearing in these cases cannot be removed by applying the renormalization procedure a finite number of times (*1*).

In this connection, one may raise the question: Is it possible to describe atomic and nuclear phenomena in terms of renormalizable interactions alone? The answer is very likely to be negative. The interaction between the electron-neutrino field and the nucleon field in Fermi's theory of beta decay is known to be, in general, a linear combination of five types of interactions. Among them, the tensor interaction, which is *not* renormalizable, is indispensable in accounting for a number of experimental results. It is not renormalizable, even if we accept the view that the beta decay is not an elementary process but can be decomposed into two stages in which creation and annihilation of a virtual meson of an unknown kind take place. Now, if the interaction between fields is not renormalizable in the ordinary sense, it amounts to the same thing to say that the procedure of renormalization necessitates the introduction without end of higher and higher derivatives of field quantities in the interaction. An interaction that involves derivatives of arbitrary order of field quantities is equivalent to a *nonlocal interaction*—that is, an interaction that refers to two or more field quantities at different space-time points. In other words, the introduction of a nonlocal interaction in field theories can be regarded as a revival of the theory of action at distance that was thought to be contradictory to the notion of field itself in classical physics. However, in the quantum theory of fields, this may not be so, because the notion of the quantized fields seems to be more flexible, field and particle being two aspects of the same physical object (*2*).

Let us consider, for example, the case of a nonlocal interaction between the scalar (or pseudoscalar) meson field $u(x_\mu)$ and the nucleon field $\psi_\alpha(x_\mu)$. The field equations can be written, in general, in the form

$$\left(\frac{\partial^2}{\partial x_\mu''\partial x_\mu''} - \kappa^2\right)u(x'') = \int \sum_{\alpha,\beta} \overline{\psi}_\alpha(x')\Phi_{\alpha\beta}(x',x'',x''')\psi_\beta(x''')\,dx'dx''', \qquad (2)$$

$$\sum_\beta\left(\gamma_\mu\frac{\partial}{\partial x_\mu'} + M\right)_{\alpha\beta}\psi_\beta(x') = -\int \sum_\beta \Phi_{\alpha\beta}(x',x'',x''')u(x'')\psi_\beta(x''')\,dx''dx''', \qquad (3)$$

where κ and M are masses of the meson and nucleon in units of $\hbar/c$, and γ_μ are Dirac matrices, $\Phi_{\alpha\beta}(x', x'', x''')$ is a matrix with four rows and columns, each matrix element being a function of three space-time points x', x'', x'''. The most general nonlocal interaction as characterized by arbitrary three-point functions $\Phi_{\alpha\beta}(x', x'', x''')$ reduces to the familiar local scalar coupling, if $u(x)$ is a scalar field and

$$\Phi_{\alpha\beta}(x',x'',x''') = g\,\delta_{\alpha\beta}\,\delta(x'-x'')\,\delta(x''-x'''),$$

where g is the coupling constant. Similarly, it reduces to local pseudoscalar coupling, if $u(x)$ is a pseudoscalar field and

$$\Phi_{\alpha\beta}(x',x'',x''') = g\,(\gamma_5)_{\alpha\beta}\,\delta(x'-x'')\,\delta(x''-x''').$$

The quantization of the fields can be carried out as usual. However, an essential departure from the local interaction theory is inevitable on account of the absence of a Schrödinger equation as such for the whole system in nonlocal interaction theory. The role of the Schrödinger equation was to determine uniquely the Schrödinger function or the probability amplitude at any time instant t in terms of the function at the immediately preceding instant $t-dt$. This was possible in the usual field theory, because the Hamiltonian $H(t)$ for the whole system depended only on the field quantities at the instant t. Once we introduce a nonlocal interaction in a relativistically invariant manner, we can no longer have a Hamiltonian that satisfies the afore-mentioned requirement. We really do not know what would be the substitute for the Schrödinger equation, or anything about any final formulation of nonlocal theories.

We know, however, that there is a formulation of ordinary field theory which seems to be suited for extension to nonlocal theories. Namely, one can define an S-matrix, which characterizes the statistical relationship between the possible results of experiments at a remote future ($t \rightarrow +\infty$) and the given results of experiments in the remote past ($t \rightarrow -\infty$), in terms of Schrödinger functions at $t=+\infty$ and $t=-\infty$ in quantum mechanics. Heisenberg pointed out that the S-matrix might well remain significant in future theories of elementary particles, whereas the Schrödinger function itself might be removed from the picture. In fact, the field equations in nonlocal interaction theory, Eqs. 2 and 3, can be integrated directly by using the same method of successive approximation as is used in ordinary field theory, which enables us to construct the S-matrix as a series in powers of the coupling constant. The trouble with local field theories with local interactions was that each term in the power series for the S-matrix was infinite because of infinite self-energies and some other infinite quantities. Recently, Møller and Kristensen have shown that, if we choose the *form function* $\Phi_{\alpha\beta}(x', x'', x''')$ in nonlocal interaction theory suitably, the self-energies of both the meson and the nucleon become finite, at least in the first approximation. In other words, the masses of these particles could be renormalized without getting into trouble of divergence. This gives us a new impetus to proceed further in this direction.

Nonlocal Fields

The introduction of nonlocal interaction between local fields was the first step toward the solution of the problem of masses of elementary particles. However, another step must be taken, if we want to approach nearer to a unified theory of elementary particles. The concept of a *nonlocal field* (*3*) was introduced in order to describe relativistically a system that was elementary in the sense that it could no longer be decomposed into more elementary constituents but was so substantial, nevertheless, as to be able to contain implicitly a great variety of particles with different masses, spins, and other intrinsic properties. For instance, a nonlocal scalar field is defined as a scalar function that depends on two sets x', x'' of space-time parameters and can be written as

$$(x_\mu' \mid \varphi \mid x_\mu'') \equiv \varphi(X_\mu, r_\mu),$$

where

$$X_\mu = \tfrac{1}{2}(x_\mu' + x_\mu''), \qquad r_\mu = x_\mu' - x_\mu''.$$

The free field equation is supposed to have the general form

$$F(\partial/\partial X_\mu, r_\mu, \partial/\partial r_\mu)\,\varphi(X_\mu, r_\mu) = 0, \qquad (4)$$

where the operator F is a certain function of $\partial/\partial X_\mu$, r_μ, and $\partial/\partial r_\mu$, which is invariant under any inhomogeneous Lorentz transformation. In particular, if we assume that F is linear in $\partial^2/\partial X_\mu\partial X_\mu$ and separable—that is,

$$F \equiv -\frac{\partial^2}{\partial X_\mu \partial X_\mu} + F^{(r)}\left(r_\mu r_\mu, \frac{\partial^2}{\partial r_\mu \partial r_\mu}, r_\mu\frac{\partial}{\partial r_\mu}\right), \qquad (5)$$

then we have eigensolutions of the form $\varphi \equiv u(X)\chi(r)$, where u and χ satisfy

$$[\partial^2/\partial x_\mu \partial x_\mu - \mu]u(X) = [F^{(r)} - \mu]\chi(r) = 0, \qquad (6)$$

μ being the separation constant. Thus, the masses of the free particles, which are associated with the nonlocal field φ, are given as the eigenvalues of the square root of the operator $F^{(r)}$ which characterizes, so to speak, the internal structure of the elementary nonlocal system. If one chooses the square root of the operator $F^{(r)}$ such that the eigenvales $\sqrt{\mu_n} = m_n$ are all positive and discrete, one can expand an arbitrary nonlocal field into a series of internal eigenfunctions $\chi_n(r)$:

$$\varphi(X, r) = \sum_n u_n(X)\chi_n(r). \qquad (7)$$

Now, when the nonlocal scalar field $(x' \mid \varphi \mid x'')$ interacts, for instance, with a local spinor field $\psi_a(x_\mu)$, the field equations become

$$\left(\frac{\partial^2}{\partial X_\mu \partial X_\mu} + F^{(r)}\right)\varphi(X, r) = -g \sum_a \psi_a\left(X + \frac{r}{2}\right)\overline{\psi_a}\left(X - \frac{r}{2}\right); \qquad (8)$$

$$\left(\gamma_\mu \frac{\partial}{\partial x_\mu'} + M\right)\psi(x') = -g\int(x' \mid \varphi \mid x'')\psi(x'')dx''. \qquad (9)$$

We insert Eq. 7 in Eq. 8, multiply both sides by the complex conjugate of χ_n, and integrate over the four-dimensional r-space, provided that $\chi_n(r)$ is quadratically integrable and, therefore, is normalized. The result is

$$\left(\frac{\partial^2}{\partial x_\mu'' \partial x_\mu''} - m_n^2\right)u_n(x_\mu'') = \int \Phi_n(x', x'', x''') \sum_a \overline{\psi}_a(x')\psi_a(x''')\, dx'dx''', \quad (10)$$

where

$$\Phi_n(x', x'', x''') \equiv g\, \tilde{\chi}_n(x' - x''')\,\delta\frac{x' + x'''}{2}\left(-x''\right). \quad (11)$$

Similarly, we obtain from Eq. 7 the equation

$$\left[\gamma_\mu \frac{\partial}{\partial x_\mu'} + M\right]\psi(x') = -\int \tilde{\Phi}_n(x', x'', x''')u_n(x'')\psi(x''')dx''dx'''. \quad (12)$$

If we compare them with the field equations, Eq. 2 and Eq. 3, of the case of nonlocal interaction between local scalar and spinor fields, we notice that the internal eigenfunction $\chi_n(r)$ characterizes the form function for the particle with mass m_n. The essential difference between the theory of nonlocal field and that of nonlocal interaction is that, in the former case, we have to take into account simultaneously all the particles with different masses m_n that are derived from an eigenvalue problem. Furthermore, the form function for each of these particles is uniquely determined by the same eigenvalue problem.

The foregoing general considerations can be illustrated by assuming a very simple form

$$F \equiv -\frac{\partial^2}{\partial X_\mu \partial X_\mu} + \frac{\lambda^2}{2}\left(-\frac{\partial^2}{\partial r_\mu \partial r_\mu} + \frac{1}{\lambda^4}r_\mu r_\mu\right)^2, \qquad (13)$$

where λ is a small constant with the dimension of length. One may call this the four-dimensional oscillator model for the elementary particle. It was considered first by Born (*4*) in connection with his idea of self-reciprocity. However, our model differs from his in that the internal structure of the particle appears explicitly in our case in connection with the nonlocalizability of the field itself. One can easily see that the mass spectrum for our case is discrete and is given by

$$m(n_1, n_2, n_3, n_0) = 2^{1/2}\lambda^{-1} \mid n_1 + n_2 + n_3 - n_0 + 1 \mid, \quad (14)$$

where n_1, n_2, n_3, and n_0 are zero or positive integers. The main trouble with the four-dimensional eigenvalue problems is the infinite degeneracy. The theory will be necessarily more complicated, if we try to get rid of this difficulty. In any case, what has been discussed here (*5*) is just the beginning of an attempt, which may lead us to a possible formulation of a unified theory of elementary particles, if we are lucky.

In conclusion, let me say that there are a number of important points that are not discussed here at all. One is the validity of the weak coupling approximations in the theories of elementary particles. We are well aware of the limitations of such approximations in connection with the problem of nuclear forces, but we cannot depart from it easily, simply because we do not have yet any satisfactory *relativistic* theory of quantized fields which is free from the assumptions of weak coupling.

References and Notes

1. Relationships between mixed field theory and nonlocal interaction theory have been discussed extensively by A. Pais and G. Uhlenbeck [*Phys. Rev.* **79**, 145 (1950)]. As for the classification of local interactions according to their renormalizability or unrenormalizability, refer to S. Sakata, H. Umezawa, and S. Kamefuchi, *Progr. Theoret. Phys. (Japan)* **7**, 377 (1950); H. Umezawa, *ibid.* **7**, 551 (1950).
2. Theories of local fields with nonlocal interactions have been discussed by many authors. Most recent of these papers are C. Bloch, *Danske Videnskab. Selskab* **27**, No. 8 (1952); P. Kristensen and C. Møller, *ibid.* No. 7.
3. H. Yukawa, *Phys. Rev.* **77**, 219 (1950); **80**, 1047 (1950); **91**, 415, 416 (1953).
4. M. Born, *Revs. Mod. Phys.* **21**, 463 (1949).
5. This article is based on a lecture given at the third meeting of Nobel prize winners in Lindau, Bodensee, 2 July 1953.

I have no economic radar to penetrate the future, but we can make it what we will it to be. Of that I am sure.—BERNARD BARUCH.