MAGNETIC MONOPOLES, FIBER BUNDLES, AND GAUGE FIELDS

Chen Ning Yang

Institute for Theoretical Physics
State University of New York at Stony Brook
Stony Brook, New York 11794

The reports in this monograph have shown great enthusiasm and exuberance for the unification of various interactions through the concept of gauge fields. I would like to emphasize a point that has not yet been explicitly stated by any of the other authors: gauge fields are deeply related to some profoundly beautiful ideas of contemporary mathematics, ideas that are the driving forces of part of the mathematics of the last 40 years. Recalling the relationship between physics and mathematics in earlier periods, general relativity and Riemannian geometry, quantum mechanics and Hilbert space, it is all too obvious that physicists may again be zeroing in on a fundamental new secret of nature.

The mathematical development referred to above is the theory of fiber bundles. It may appear, a priori, that this theory is quite abstract and is unrelated to the structure of the physical world. To show that this is not true, we will start with a simple demonstration that electromagnetism and quantum mechanics together lead naturally to "nontrivial fiber bundles." We will then trace the early history of the gauge field concept and its generalization, emphasizing three related but different conceptual motivations, each of which leads to a general formulation of gauge fields.

MAGNETIC MONOPOLES AND NONTRIVIAL BUNDLES

The magnetic monopole is the magnetic charge. Though the idea of magnetic monopoles probably was discussed in classic physics early in the history of electricity and magnetism, modern discussions of this concept date back only to 1931, when the important paper of Dirac[1] pointed out that magnetic monopoles in quantum mechanics exhibit some extra and subtle features. In particular, with the existence of a magnetic monopole of strength g, electric charges and magnetic charges must necessarily be quantized, in quantum mechanics. We will give a new derivation of this result below.

If one wants to describe the wave function of an electron in the field of a magnetic monopole, it is necessary to find the vector potential **A** around the monopole. Dirac chose a vector potential that has a string of singularities. The necessity of such a string of singularities is obvious if we prove the following theorem[2]:

Theorem: Consider a magnetic monopole of strength $g \neq 0$ at the origin, and consider a sphere of radius R around the origin. There does not exist a vector potential **A** for the monopole magnetic field that is singularity free on the sphere.

This theorem can be proved easily in the following way. If there were a singularity-free **A**, we would consider the loop integral

$$\oint A_\mu dx^\mu$$

around a parallel on the sphere, as indicated in FIGURE 1. According to Stoke's theorem, this loop integral is equal to the total magnetic flux through the cap α:

$$\oint A_\mu dx^\mu = \Omega_\alpha. \tag{1}$$

Similarly, we can apply Stoke's theorem to cap β, obtaining

$$\oint A_\mu dx^\mu = \Omega_\beta. \tag{2}$$

Here, Ω_α and Ω_β are the total upward magnetic fluxes through caps α and β, both of which are bordered by the parallel. Subtracting these two equations, we obtain

$$0 = \Omega_A - \Omega_B, \tag{3}$$

which is equal to the total flux *out* of the sphere, which, in turn, is equal to $4\pi g \neq 0$. We have thus reached a contradiction.

Having proved this theorem, we observe that R is arbitrary. Thus, one concludes that there must be a string(s) of singularities in the vector potential to describe the monopole field. Yet, we know that the magnetic field around the monopole is singularity free. This fact suggests that the string of singularities is not a real physical difficulty. Indeed, the situation is reminiscent of the problem that one faces when one wants to find a parametrization of the surface of the globe. The coordinate system that we usually use, latitude and longitude, is not singularity free. It has singularities at the north pole and at the south pole. Yet, the surface of the globe is evidently devoid of singularities. We deal with this situation usually in the manner illustrated in FIGURE 2. We consider a rubber sheet with nicely defined coordinates and stretch and wrap it downward onto the globe, so that it covers more than the northern hemisphere. Similarly, we consider another rubber sheet with nicely defined coordinates and stretch and wrap it upward, so it covers more than

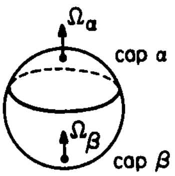

FIGURE 1. A sphere of radius R with a magnetic monopole at its center. The parallel divides the sphere into two caps α and β.

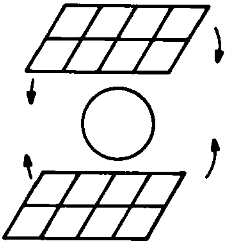

FIGURE 2. Method of parametrizing the globe.

the southern hemisphere. We now have a double system of coordinates to describe the points on the globe. The description is analytic in the domain covered by each sheet, if the globe had experienced no violence in the stretching and wrapping. In the overlapping region covered by both sheets, one has two coordinate systems that are transformable into each other by an analytic nonvanishing Jacobian. This double coordinate system is an entirely satisfactory way to parametrize the globe.

Following this idea, we will now attempt to exorcise the string of singularities in the monopole problem by dividing space into two regions. We will call the points outside of the origin, above the lower cone in FIGURE 3, region R_a. Similarly, we will call the points outside of the origin, under the upper cone, R_b, The union of these two regions gives all points outside of the origin. In R_a, we will choose a vector potential for which there is only one nonvanishing component of A, the azimuthal component:

$$(A_r)_a = (A_\theta)_a = 0, \qquad (A_\phi)_a = \frac{g}{r\sin\theta}(1-\cos\theta), \tag{4}$$

It is important to notice that this vector potential has no singularities anywhere in R_a. Similarly, in R_b, we choose the vector potential

$$(A_r)_b = (A_\theta)_a = 0 \qquad (A_\phi)_b = \frac{-g}{r\sin\theta}(1\times\cos\theta), \tag{5}$$

which has no singularities in R_b. It is simple to prove that the curl of either of these two potentials gives correctly the magnetic field of the monopole.

In the region of overlap, because both of the two sets of vector potentials share the same curl, the difference between them must be curlless and therefore must be a gradient. Indeed, a simple calculation shows

$$(A_\mu)_a - (A_\mu)_b = \partial_\mu \alpha, \quad \text{where } \alpha = 2g\phi, \tag{6}$$

where ϕ is the azimuthal angle. The Schrödinger equation for an electron in the monopole field is thus

$$\frac{1}{2m}(p-eA_a)^2\psi_a + V\psi_a = E\psi_a, \quad \text{in } R_a,$$

$$\frac{1}{2m}(p-eA_b)^2\psi_b + V\psi_b = E\psi_b, \quad \text{in } R_b,$$

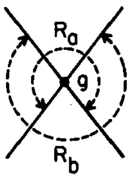

FIGURE 3. Division of space outside of monopole g into overlapping regions R_a and R_b.

where ψ_a and ψ_b are, respectively, the wave functions in the two regions. The fact that the two vector potentials in these two equations are different by a gradient tells us, by the well-known gauge principle, that ψ_a and ψ_b are related by a phase factor transformation

$$\psi_a = S\psi_b, \quad S = \exp(ie\alpha), \tag{7}$$

or

$$\psi_a = [\exp(2iq\phi)]\psi_b, q = eg. \tag{8}$$

Around the equator, which is entirely in R_a, ψ_a is single valued. Similarly, because the equator is also entirely in R_b, ψ_b is single valued around the equator. Therefore, S must return to its original value when one goes around the equator. That fact implies Dirac's quantization condition:

$$2q = \text{integer}. \tag{9}$$

Hilbert Space of Sections

Two ψs, ψ_a and ψ_b, in R_a and R_b, respectively, that satisfy the condition of transition (Equation 8) in the overlap region are called a *section* by the mathematicians. We see that around a monopole, the electron wave function is a *section* and *not an ordinary function*. We will call these functions wave sections.

Different wave sections (which belong to different energies, for example) clearly satisfy the same condition of transition (Equation 8) with the same q. Thus, we need to develop [3] the concept of a Hilbert space of sections. To develop this concept, we define the scalar product of two sections ξ, η (for the *same* q) by

$$(\eta, \xi) = \int \eta^* \xi d^3 r. \tag{10}$$

(The question of convergence at $r = 0$ and $r = \infty$ is ignored here.) Notice that in the overlap

$$(\eta_a)^* \xi_a = (\eta_b)^* \xi_b, \tag{11}$$

so that Equation 10 is well defined.

It is clear that if ξ is a section, $x\xi$ is also a section, because

$$x\xi_a = S(x\xi_b).$$

Thus, x is an *operator* in the Hilbert space of sections. Similarly, we prove that the components of $(\mathbf{p} - e\mathbf{A})$ are operators, but those of $\mathbf{p}$ are not. Furthermore, $\mathbf{x}$ and $\mathbf{p} - e\mathbf{A}$ are both Hermitian.

Following Fierz,[4] we will now attempt to construct angular momentum operators. Define

$$\mathbf{L} = \mathbf{r} \times (\mathbf{p} - e\mathbf{A}) - \frac{q\mathbf{r}}{r}. \tag{12}$$

It is clear that L_x, L_y, and L_z are Hermitian operators on the Hilbert space of sections. The following commutation rules can be easily verified:

$$\begin{aligned} &[L_x, x] = 0, \quad [L_x, y] = iz, \quad [L_x, z] = -iy, \\ &[L_x, p_x - eA_x] = 0, \quad [L_x, p_y - eA_y] = i(p_z - eA_z), \\ &[L_x, p_z - eA_z] = -i(p_y - eA_y). \end{aligned} \tag{13}$$

It follows from these commutation rules that

$$[L_x, L_y] = iL_z, \text{ etc.} \tag{14}$$

Equation **13**, together with its consequence (Equation **14**), show that L_x, L_y, and L_z are the *angular momentum operators.*[4] We emphasize that neither the Hilbert space nor these operators possess any "singularities." (The singularities of A_a and A_b are not real singularities, because they occur outside of R_a and R_b, respectively.)

Monopole Harmonics $Y_{q,l,m}$

Because $[r^2, \mathbf{L}] = 0$, we can diagonalize r^2 and study operators $\mathbf{L}$ for fixed r^2. That is, we will study sections of the form

$$\delta(r^2 - r_0^2)\xi,$$

where ξ is a section dependent only on angular coordinates θ and ϕ. $\mathbf{L}$ operates, then, on "angular sections."

Equation **14** shows that $[L^2, L_z] = 0$. Simultaneous diagonalization produces the familiar multiplets with eigenvalues $l(l+1)$ and m,

$$L^2 Y_{q,l,m} = l(l+1) Y_{q,l,m}; \; L_z Y_{q,l,m} = m Y_{q,l,m}, \tag{15}$$

where $l = 0, 1/2, 1, \ldots$, and for each value of l,m ranges from $-l$ to $+l$ in integral steps of increment. $Y_{q,l,m}$ are eigensections, which are called[3] monopole harmonics. The allowed values of l and m are

$$l = |q|, \; |q| + 1, \; |q| + 2, \ldots, \qquad m = -l, \; -l+1, \ldots, l. \tag{16}$$

Each of these l,m combinations occurs exactly once. One can choose each Y normalized, so that

$$\int_0^\pi \sin\theta d\theta \int_0^{2\pi} |Y_{q,l,m}|^2 d\phi = 1. \tag{17}$$

Different $Y_{q,l,m}$ (for fixed q) are orthogonal, a fact one easily proves in the usual way from Equation **15**.

The explicit values of $Y_{q,l,m}$ in terms of Jacobi polynomials were given in Reference 3. They were obtained from Equation **15**, in exactly the same way one usually obtains the spherical harmonics $Y_{l,m}$. Indeed,

$$Y_{l,m} = Y_{0,l,m}.$$

The collection of $Y_{q,l,m}$ for fixed q and values of l,m given by Equation **16** form[3] a complete orthonormal set of angular sections.

Each $(Y_{q,l,m})_a$ is analytic in R_a; so is $(Y_{q,l,m})_b$ in R_b. Thus, all of the discontinuities, cusps, and singularities in **A** and in ψ are removed in a very smooth way.

Remarks: (A) It is important to realize that the above-described way of using $(A)_a$ and $(A)_b$ together to describe the magnetic field of a monopole has an additional advantage: It gives the magnetic field **H** correctly *everywhere*. In older papers, one often used a single **A** with a string of singularities. Because, by definition,

$$\nabla \cdot (\nabla \times A) = 0,$$

the magnetic field described by $\nabla \times A$ must have *continuous* flux lines. Thus, its flux lines consist of the dotted lines of FIGURE 4, plus the bundle of lines described by the solid line, so as to make the net flux at the origin zero. Thus, $\nabla \times A$ does not correctly describe the magnetic field of the monopole, a point already emphasized by Wentzel.[5]

(B) For ordinary spherical harmonics, there are many important theorems, such as the spherical harmonics addition theorem and the decomposition of products of spherical harmonics by use of Clebsch-Gordon coefficients. These theorems can be generalized to monopole harmonics.[6]

(C) In the approximately 40 years since Dirac's first paper on monopoles, the subject has been beset with difficulties due to singularities. Now that we have removed the difficulty of string singularities through the introduction of the concept of sections, it is revealed that there is yet another difficulty, which we will call the Lipkin-Weisberger-Peshkin[7] difficulty. This difficulty occurs[8] in studying the radial wave function of a Dirac electron around a monopole (TABLE 1). It can be removed through the introduction of a small extra magnetic moment for the Dirac electron.

(D) It is instructive to go back to the reasoning represented in FIGURE 1 and attempt to repeat the steps for the combined A_a, A_b description of the magnetic field. Choose the parallel to be the equator. Then,

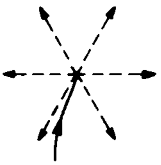

FIGURE 4. Magnetic flux lines due to **A**. Because $\nabla \cdot (\nabla \times A) = 0$, flux lines are everywhere continuous. Therefore, there is "return flux" along the solid line.

TABLE 1

DIFFICULTIES AND METHODS OF SOLUTION FOR STUDYING THE MOTION OF A DIRAC ELECTRON IN THE FIELD OF A MAGNETIC MONOPOLE

Angular Wave Function	Radial Wave Function
Difficulty of string singularity, solved by introducing sections	Lipkin-Weisberger-Peshkin difficulty, solved by introducing extra magnetic moment

$$\oint (A_\mu)_a \mathrm{d}x^\mu = \Omega_\alpha,$$

$$\oint (A_\mu)_b \mathrm{d}x^b = \Omega_\beta.$$

Thus,
$$4\pi g = \Omega_\alpha - \Omega_\beta = \oint [(A_\mu)_\alpha - (A_\mu)_\beta] \mathrm{d}x^\mu,$$

which is, by Equation 6, equal to the increment of α around the equation, that is, $2g(2\pi) = 4\pi g$.

We have arrived at an identity. I have provided this simple argument because it is exactly the gist of the proof of the famous Gauss-Bonnet-Allendoerfer-Weil-Chern theorem and the later Chern-Weil theorem, which play seminal roles in contemporary mathematics.

In fact, gauge fields, of which electromagnetism is the simplest example, are conceptually identical to some mathematical concepts in fiber bundle theory. TABLE 2 gives [2] translations for the terminologies used by physicists, on the one hand, and mathematicians, on the other. We notice that, in particular, Dirac's monopole quantization (Equation 9) is identical to the mathematical concept of classification of U(1) bundles according to the first Chern class.

The last two entries of TABLE 2 identify electromagnetism with and without magnetic monopoles with connections to trivial and nontrivial U(1) bundles. Why is electromagnetism without monopoles "trivial"? We can gain some understanding by looking at a paper loop and a Moebius strip (FIGURE 5). If they are cut along the dotted lines, each would break into two pieces. Looking at the resultant pieces, we cannot differentiate between the two. The paper loop and the Moebius strip are different only in the way the resultant pieces are put together. For the latter, a twist of one of the resultant pieces is necessary. The difference between a trivial and a nontrivial bundle resides only in the processes of *joining:* for the nontrivial bundle, a twist is needed in the joining process. In the case of electromagnetism, the joining process is given by Equation 7 or 8. If there is no monopole, $S = 1$, and the bundle is trivial. If there is a monopole, $S \neq 1$, and the bundle is nontrivial. (We may describe the nontrivial nature by saying that a *twist of phase* is necessary.)

EARLY HISTORY OF THE CONCEPT OF GAUGE FIELDS

Einstein's discovery of the relationship between gravitation and the geometry of space-time stimulated work by many great geometers: Levi-Civita, Cartan, Weyl, and others. In his book, *Raum, Zeit und Materie* (space, time and matter), Weyl[9]

TABLE 2
TRANSLATION OF TERMINOLOGIES

Gauge Field Terminology	Bundle Terminology
Gauge (or global gauge)	principal coordinate bundle
Gauge type	principal fiber bundle
Gauge potential b^k_μ	connection on a principal fiber bundle
S (Equation 8)	transition function
Phase factor Φ_{QP}	parallel displacement
Field strength $f^k_{\mu\nu}$	curvature
Source (electric) J^k_μ	?
Electromagnetism	connection on a U_1 bundle
Isotopic spin gauge field	connection on a SU_2 bundle
Dirac's monopole quantization	classification of U_1 bundle according to first Chern class
Electromagnetism without monopole	connection on a trivial U_1 bundle
Electromagnetism with monopole	connection on a nontrivial U_1 bundle

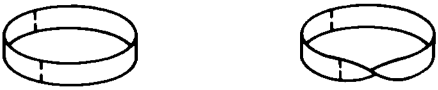

FIGURE 5. Examples of trivial (*left*) and nontrivial (or Moebius strips, *right*) fiber bundles.

attempted to unify gravity and electromagnetism through the use of the geometric concept of a space-time-dependent scale change. The basic idea is summarized below.

	$\bullet$	$\xrightarrow{\mathrm{d}x\mu}\ \bullet$
scale	1	$1 + S_\mu \mathrm{d}x^\mu$
f	f	$f + (\partial f/\partial x^\mu)\,\mathrm{d}x^\mu$
scale change	f	$f + (\partial/\partial x^\mu + S_\mu) f \mathrm{d}x^\mu$

In the summary above, the first line indicates how the scale changes in going from a point x^μ to a neighboring point $x^\mu + \mathrm{d}x^\mu$ of space-time. The second line shows how a function of space-time changes as a result of the change in argument from x^μ to $x^\mu + \mathrm{d}x^\mu$. Finally, if the scale change is applied to the function f, one obtains at $x^\mu + \mathrm{d}x^\mu$ the product

$$(f + \partial f/\partial x^\mu \mathrm{d}x^\mu) \quad (1 + S_\mu \mathrm{d}x^\mu).$$

Expanding to first order in the small displacement gives the last line in the summary. The increment in f is, then,

$$(\partial/\partial x^\mu + S_\mu) f \mathrm{d}x^\mu. \qquad (18)$$

Weyl tried to incorporate electromagnetism into a geometric theory by identifying the vector potential A_μ with a space-time-dependent S_μ, generating scale changes as described. This attempt proved, however, unsuccessful.

In 1925, the concepts of quantum mechanics emerged. A key concept in quantum mechanics is the replacement of the momentum p_μ in the classic Hamiltonian by an operator:

$$p_\mu \rightarrow -ih(\partial/\partial x^\mu).$$

For a charged particle, the replacement is

$$p_\mu-(e/c)A_\mu \rightarrow -ih[\partial/\partial x^\mu - i(e/hc)A_\mu]. \quad \textbf{(19)}$$

In 1927, Fock[10] observed that one could base quantum electrodynamics on this operator. London[11] pointed out the similarity of Fock's to Weyl's earlier work. Comparing Equations **18** and **19**, Weyl's identification would be correct if one makes the replacement

$$S_\mu \rightarrow -i(e/hc)A_\mu.$$

In other words, instead of a *scale change*

$$(1+S_\mu \mathrm{d}x^\mu),$$

one considers a *phase change*

$$[1-i(e/hc)A_\mu \mathrm{d}x^\mu] \simeq \exp[-i(e/hc)A_\mu \mathrm{d}x^\mu], \quad \textbf{(20)}$$

which can be thought of as an *imaginary scale change.* Weyl put all of these expressions together[12] in a remarkable paper (which also first discussed the two-component theory of a spin-1/2 particle) in which the transformation of the electromagnetic potential

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \alpha \text{ (second-type transformation)}, \quad \textbf{(21)}$$

and the associated phase transformation

$$\psi \rightarrow \psi' = \psi \exp(ie\alpha/hc) \text{ (first-type transformation)}, \quad \textbf{(22)}$$

of the wave function of a charged particle were explicitly discussed.[13]

Although the phase change factor (Equation **20**) is no longer a scale factor, Weyl kept the earlier terminology[*][†] that he used in 1918–20 and called both the transformation (Equation **20**) and the associated phase change of wave functions "gauge" transformations.

Generalization: With the discovery of many new particles after World War II, physicists explored various couplings between the "elementary particles." Many possible couplings can be written down, and the desire to find ***a principle to choose among the many possibilities*** was one of the motivations[17,18] for an attempt to generalize Weyl's gauge principle for electromagnetism. The point here is that for electromagnetism, the gauge principle determines, all at once, the way in which ***any*** particle of charge *qe*, a ***conserved*** quantity, serves as a ***source*** of the electromagnetic field. Because the isotopic spin **I** is also conserved, a natural question was, "Does there exist a generalized gauge principle that determines the way in which **I** serves as the source of a new field?"

Another motivation for an attempt at generalization is the observation that the conservation of **I** implies that the proton and the neutron are similar. Which to call a proton or, indeed, which superposition of the two to call a proton, is a convention that one can select arbitrarily (if the electromagnetic interaction is switched off). If one requires this freedom of choice to be independent for observers at different space-time points, that is, if one requires ***localized*** freedom of choice, one is led to a generalization of the gauge principle.

These two motivations were, of course, intertwined and led quite naturally to the formulation[18] of non-Abelian gauge fields.

A third approach[19] to a generalized gauge principle came later and is the "integral formalism" of gauge fields. It starts from the observation that the gauge principle of Weyl deals with a phase factor (Equation **20**) between two neighboring points. Along a path from space-time point A to space time point B, the resultant phase factor is

$$\Phi_{\mathrm{BA}} = \exp[-i(e/hc)\int_{\mathrm{A}}^{\mathrm{B}} A_{\mu}\mathrm{d}x^{\mu}], \qquad \textbf{(23)}$$

which is path dependent, that is, nonintegrable. (Dirac[1] had already discussed, in 1931, "non-integrable phases for wave functions.") If one analyzes the meaning of electromagnetism in quantum mechanics, especially through a discussion of the Bohm-Aharonov experiment,[20‡] one reaches the conclusion[2] that "electromagnetism is the gauge invariant manifestation of a non-integrable phase factor."

Once this conclusion is reached, a natural generalization is to replace a

[*] The idea of scale invariance, discussed in Reference 9, was developed earlier, in 1918–19, in three papers by Weyl (submitted on May 2 and June 8, 1918 and on January 7, 1919). In the first two of them, he used the term ***Massstab Invarianz*** (see Reference 14); in the third paper, he settled on the term ***Eich Invarianz***.

The English translation of ***Eich Invarianz*** was "calibration invariance" in Henry Brose's 1921 translation of the fourth edition of Weyl's book ***Space, Time and Matter***[15] (republished by Dover). The translation "gauge invariance" was not used, I suspect, until after Weyl's 1929 article.[12] It appeared (probably not for the first time) in Dirac's article[1] of 1931.

[†] The transformation (Equation **21**) that leaves field strengths unchanged must have been known in the nineteenth century. It did not, however, seem to have a specific name. In the many editions of Foppl-Abraham-Becker-Sauter on electricity and magnetism, which started in 1894, *Eich* or "gauge" was not used until the 1964 English translation ***Electromagnetic Fields and Interactions,***[16] in which the term "Lorentz gauge" was inserted in a footnote.

[‡] The experiment was performed by Chambers.[21]

"nonintegrable phase factor" by a "nonintegrable element of a Lie group." One thus obtains naturally an integral formalism of gauge fields.

We illustrate in FIGURE 6 the three approaches to the general concept of gauge fields. The three approaches are, of course, deeply interrelated, because phases, symmetry, and conservation laws are themselves related.

It is my opinion that, conceptually, the integral formalism of gauge fields is to be preferred to the earlier differential approach. The integral formalism has more structure and more meaning. It brings to the fore problems of global topology not easily formulated in terms of the differential approach. For example, in our earlier discussion of the field around the magnetic monopole, we did not introduce the concept of nonintegrable phase factors. We did not run into any conceptual difficulties, only because we had not raised such questions as a rotation of the coordinate axes. As soon as such questions are raised, it becomes apparent that the integral formalism is more superior, because it specifies that intrinsic meaning is unrelated to the choice of coordinate axes and of regions R_a and R_b.

Differential formalism, however, is used in computing. (The relationship between differential and integral formalisms is quite similar to that between Lie algebras and Lie groups.) In fact, a gauge-Riemannian calculus has been developed.[21]

Electromagnetism is, as we have seen, a gauge field. That gravitation is a gauge field is universally accepted, although exactly how it is a gauge field is a matter still to be clarified.[19,22] Whether weak and strong interactions are also due to gauge fields is a matter that has been intensively studied in recent years,[23] together with the question of the renormalizability of non-Abelian gauge fields.[24§] If one may borrow a term used by the biologists, one would say that there is gradually forming a "dogma" that all interactions are due to gauge fields. Because of the mathematical difficulties involved in the solution of quantized gauge fields,

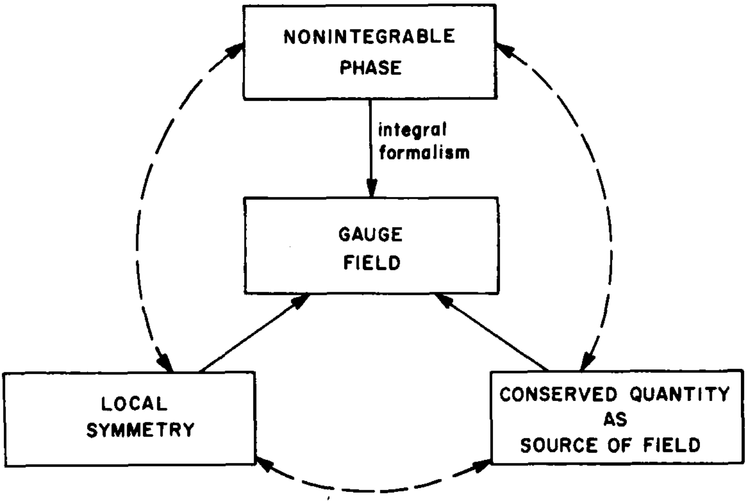

FIGURE 6. Three motivations that led to the concept of gauge fields.

§ Abers and Lee[23] also contains a review of earlier works of R. P. Feynman, L. D. Faddeev, V. N. Popov, and M. T. Veltman.

however, I believe it will be a long time before the question can be definitively answered as to exactly how strong and weak interactions are due to gauge fields.

Reflecting on how the concepts basic to gauge fields were formulated by physicists, we see that at every step, the development was tied to the problem of the conceptual description of the physical world. Firstly, Maxwell equations originated with the four fundamental experimental laws of electricity and magnetism and with Faraday's introduction of the concepts of field and flux. Maxwell's equations and the principles of quantum mechanics led to the idea of gauge invariance. Attempts to generalize this idea, motivated by physical concepts of phases, symmetry, and conservation laws, led to the theory of non-Abelian gauge fields. That non-Abelian gauge fields are conceptually identical to ideas in the beautiful theory of fiber bundles, developed by mathematicians *without reference to the physical world,* was a great marvel to me. In 1975, I discussed my feelings with Chern, and said, "This is both thrilling and puzzling, since you mathematicians dreamed up these concepts out of nowhere." He immediately protested, "No, no, these concepts were not dreamed up. They were natural and real."

References

1. Dirac, P.A.M. 1931. Proc. Roy. Soc. **A133:** 60.
2. Wu, T. T. & C. N. Yang. 1975. Phys. Rev. D **12:** 3845.
3. Wu, T. T. & C. N. Yang. 1976. Nucl. Phys. B **107:** 365.
4. Fierz, M. 1944. Helv. Phys. Acta **17:** 27.
5. Wentzel, G. 1966. Progr. Theor. Phys. Suppl. **37–38:** 163.
6. Wu, T. T. & C. N. Yang. 1977. To be published.
7. Lipkin, H. J., W. I. Weisberger & M. Peshkin. 1969. Ann. Phys. **53:** 203.
8. Kazama, Y., C. N. Yang & A. S. Goldhaber. 1977. Phys. Rev. D. In press.
9. Weyl, H. 1920. Raum, Zeit und Materie. 3rd edit. Springer Verlag. Berlin-Heidelberg. New York.
10. Fock, V. 1927. Z. Phys. **39:** 226.
11. London, F. 1927. Z. Phys. **42:** 375.
12. Weyl, H. 1929. Z. Phys. **56:** 330.
13. Pauli, W. 1933. Handbuch der Physik. 2nd edit. Vol. 24(1): 83. Geiger and Scheel.; Pauli, W. 1941. Rev. Mod. Phys. **13:** 203.
14. Weyl, H. 1918. Sitzber. Preuss Akad. Wiss.: 465; Weyl, H. 1918. Math. Z. **2:** 384; Weyl, H. 1919. Ann. Phys. **59:** 101.
15. Weyl, H. 1921. Space, Time and Matter. Dover Publications, Inc. New York, N.Y.
16. 1964. Electromagnetic Fields and Interactions. Blaisdell Publishing Co. Waltham, Mass.
17. Yang, C. N. & R. Mills. 1954. Phys. Rev. **95:** 631.
18. Yang, C. N. & R. Mills. 1954. Phys. Rev. **96:** 191.
19. Yang, C. N. 1974. Phys. Rev. Lett. **33:** 445.
20. Aharonov, Y. & D. Bohm. 1959. Phys. Rev. **115:** 485; Chambers, R. G. 1960. Phys. Rev. Lett. **5:** 3.
21. Yang, C. N. 1975. Proc. Sixth Hawaii Topical Conf. Particle Phys.
22. Utiyama, R. 1956. Phys. Rev. **101:** 1957.
23. Weinberg, S. 1967. Phys. Rev. Lett. **19:** 1264; Salam, A. 1968. *In* Elementary Particle Theory. N. Svartholm, Ed. Almquist and Forlag. Stockholm, Sweden.
24. 'tHooft, G. 1971. Nucl. Phys. B **35:** 167; Abers, E. S. & B. W. Lee. 1973. Phys. Rep. **9C:** 1.