Light as a fundamental particle

The question of whether the photon is "special" leads to some remarkable conclusions about the interactions of matter and about the underlying symmetry of nature.

Steven Weinberg

We take it pretty much for granted that the whole visible world of matter and radiation can be explained, if not in fact at least in principle, in terms of the interactions of a handful of so-called "elementary particles": the electron; the proton; the neutron; the quantum of light, the photon; the quantum of gravitational radiation, the graviton, and perhaps also the neutrino. We would like to know why these particles have the properties they have, and therefore why the world is the way it is. Or, if you do not believe that scientists should ask "why," you can restate the question in this form: What we want to know is the set of simple principles from which the properties of these particles, and hence everything else, can be deduced.

In our search for these principles we have uncovered a large zoo of other particles, some of which are listed in Table 1. The particles that we now include in the table of elementary particles are of varying familiarity. Among them, of course, are the photon, the graviton, the neutrino, the electron, the proton and the neutron. But there are other members of the list—the muon, the pi meson, the K meson—that are less familiar. Some of these, in fact, can be created only by cosmic rays or by artificial beams in accelerators. These various particles are distinguished from one another according to their masses, spins, charges and other properties, but the question of how familiar the particles are is basically a matter of their lifetime. The particles that are most familiar are naturally those of long lifetimes—photons, gravitons, electrons, neutrinos and protons, as well as neutrons—which are stable in nuclei although not in free space.

As you go to more and more unstable particles, they naturally become less and less familiar. But we really see nothing in this table to indicate that any one particle is fundamental in a way that other particles are not. This article originated as a talk at a meeting of the Optical Society of America while we know that there is no "Muon Society of America." The muon, of course, plays a much smaller role in everyday life than the photon but that is an accident. The photon happens to be stable while the muon is not, but we do not see any reason to suppose that the muon is in any sense a less fundamental particle than the photon—or any of the others.

But is that really true? Is it true that the photon is just another particle, distinguished from the others by a particular value of charge, spin, mass, lifetime and certain interaction properties? Or is there really something special about photons? Do they play in some sense a fundamental role; have they a deeper relation to the ultimate formulas of physics than the other particles?

I can—and will—argue both sides of this issue with great conviction. I will first present the case for the hypothesis that the photon is just another particle and that its properties—in fact, the whole of electrodynamics and optics—can be understood as flowing very simply from its particle properties, especially mass and spin, as given in Table 1. Then I will take the other side, and show why I think the photon really expresses something fundamental about the laws of nature. (If such a dialog appears to be rather confusing, I am sorry.) This "something fundamental" is symbolized in figure 1. The Crab nebula is photographed in visible light, but the supernova that produced it may, as we shall see, have been caused by forces related to, but less familiar than, electromagnetism.

Steven Weinberg is Higgins Professor of Physics at Harvard University and Senior Scientist at the Smithsonian Astrophysical Observatory. He gave an invited paper at the Washington meeting of the Optical Society of America, Spring 1974, and this article is an adaptation of a tape recording of that talk.

Light as a particle

The peculiar properties of the photon that allow us to go so far in deducing all of electrodynamics and optics on an *a priori* basis arise from the fact that it is a mass-zero particle with integer spin. In a 1939 paper, Eugene Wigner first explained how to analyze the states of a particle with a definite spin in the context of special relativity and quantum mechanics, in terms of the so-called "little group." This is the subgroup of the Lorentz group that leaves the state of motion unchanged. The little group is, very simply, the set of all the Lorentz transformations that do not change the velocity of the particle.

If we have a massive particle at rest, the little group is very well known: It just consists of all the rotations—obviously, the only Lorentz transformations that do not give motion to a particle at rest are rotations. These rotations are indicated schematically in figure 2 by small circles going around the x-axis, the y-axis and the z-axis. If we imagine the particle being given a velocity, say upwards along the z-axis, the Lorentz transformation, although it will do nothing to the little circles in the xy plane around the z-axis, will flatten out the circles around the x-axis and around the y-axis. As we go to higher and higher velocities, these circles become straight lines.

And so the little group, in the limiting case of a particle moving at the speed of light (as any massless particle does) is no longer the rotation group but something else: the product of a

HALE OBSERVATORIES

Visible light made this photo of the planetary nebula in Taurus, the Crab; but the supernova that created it in the year 1054 may well have involved other interactions. Despite their differences, do the basic forces of physics share similar invariances? Figure 1

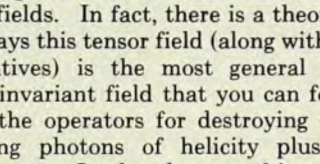
simple set of rotations only around the z-axis with the set of translations on the xy plane. Now it is precisely the fact that there is not only a J_z but also a J_x and a J_y—not only a component of angular momentum around the direction of motion, but also components perpendicular to it—that leads us to the famous conclusion of quantum mechanics that a particle with spin one and mass not equal to zero can exist in three states, with a z-component of spin equal to plus 1, zero and minus 1: three polarization states.

However, we do not have this conclusion for a particle moving at the speed of light, for the simple reason that the little group is not the rotation group. It only has one rotation, J_z. From a completely general point of view there is no reason why a particle with $J_z = +1$ has to be accompanied by a particle with J_z equal to zero or to any other value. In fact, parity conservation does require that it be accompanied by a particle with $J_z = -1$, and so we come to the conclusion that a particle with unit spin that moves at the speed of light has to exist in *two* polarization states, characterized by helicity +1 and −1, depending on whether its spin vector is parallel or antiparallel with its momentum. In optics, these states are known more familiarly as states of right and left circular polarization. Thus we arrive, without knowing anything about Maxwell's equations, at the first fact about photons: that they exist in only two polarization states. (This is necessary for my argument because I am trying to get Maxwell's equations!)

From the operators that create or destroy photons in these two polarization states one can attempt to construct various kinds of quantum fields, characterized by diverse Lorentz-transformation properties. One familiar example is the antisymmetric tensor formed from the six components of the electric and magnetic fields. In fact, there is a theorem that says this tensor field (along with its derivatives) is the most general Lorentz-invariant field that you can form from the operators for destroying and creating photons of helicity plus or minus one. In the absence of interactions with other particles or fields, the electric and magnetic fields formed in this way automatically satisfy Maxwell's free-field equations. Nothing could be more satisfactory.

Interactions

Now, what about interactions? If I imagine charged particles interacting by transmitting this kind of photon, and if I use the field-strength tensor mentioned in the previous paragraph to describe the interaction of the photon field with the charge, I will find that I get a factor of $\sqrt{p}$, where p is the momentum carried by the photon, at both vertices of the Feynman diagram:



The propagation of the photon introduces a factor of $1/p^2$, and multiplying $\sqrt{p}$, $1/p^2$ and $\sqrt{p}$ again, I get $1/p$, which under a Fourier transformation gives $1/r^3$, the familiar dipole–dipole potential. That is, if we use only field strengths to describe the interactions of electromagnetic waves and charged particles, then we will find that the interactions between photons and other particles generate only $1/r^3$ potentials rather than the famous long-range Coulomb potential, $1/r$.

How does the $1/r$ potential come into physics? Well, as everyone knows, it comes into physics because the interaction of photons with charged particles takes place not only through the field-strength tensor, but also through the vector potential, the curl of which is the field-strength tensor. This statement appears to contradict the theorem I mentioned above, that says that the only fields that are formed from the creation and annihilation of photons with helicity plus or minus one are tensor fields and their derivatives, because here we seem to have a vector field!

However, if you remember that there are only two polarization states, you realize that the vector potential A_μ is *not* a vector. In fact, if you calculate how it transforms under a Lorentz transformation, you find that it picks up an additional gradient term $\partial\phi/\partial x^\mu$. This may be less surprising if I remind you that the only vector potential that could actually be formed in this way, without introducing spurious degrees of freedom, is what is usually called the "Coulomb-gauge" or "radiation-gauge" vector potential, defined to satisfy certain non-Lorentz-invariant constraints, such as, for instance, that the time component of the potential should vanish.

So then how can the Lorentz invariance be satisfied? It will only be satisfied if this vector potential interacts with a current J^μ in such a way that the extra gradient term in the Lorentz transformation law does not matter. In other words, the vector potential must interact with currents in such a way that the current is a four-vector and the divergence of the current is zero, so that when you integrate by parts, you find that the integral of $J^\mu\partial\phi/\partial x^\mu$ is zero.

Therefore we conclude that the only way of describing particles of mass zero and spin one that will give rise to the long-range forces, that is, $1/r^2$ forces, derived from $1/r$ potentials, is to use a vector potential that is not a four-vector; one, therefore, that requires the interaction with a conserved current—not just any old vector current!

Gauge invariance

Now, there is another way of saying all this, and that is that the field equations must be unchanged under a *gauge transformation*. This is a phase transformation in which the change in phase of any field is proportional to the charge q that is destroyed by that field times $\phi(x)$, an arbitrary function of space-time:

$$\psi(x) \longrightarrow e^{iq\phi(x)}\psi(x) \qquad (1)$$

In this gauge transformation, the vector potential undergoes a translation by the amount $\partial\phi(x)/\partial x^\mu$. The fact that A_μ couples to a conserved current means that this transformation will not in fact change the field equations.

This gauge invariance, then, which leads inevitably to Maxwell's equations, is a consequence of the requirements of Lorentz invariance, as applied to particles of mass zero and spin one. And with this we have the whole formalism of electrodynamics and optics: It follows inevitably if you believe that photons exist, of mass zero and spin one.

There is an interesting analogy between gauge invariance on the one hand and the general covariance of gravitational theory on the other. In fact, this whole line of argument could be repeated, starting with the idea that there exists a fundamental particle, the graviton, with spin two and mass zero. It may sound a little artificial to do it that way, because no one has detected any gravitons yet. There are, however, good reasons why no one has, and I think this is a perfectly defensible line of argument.

Turnabout

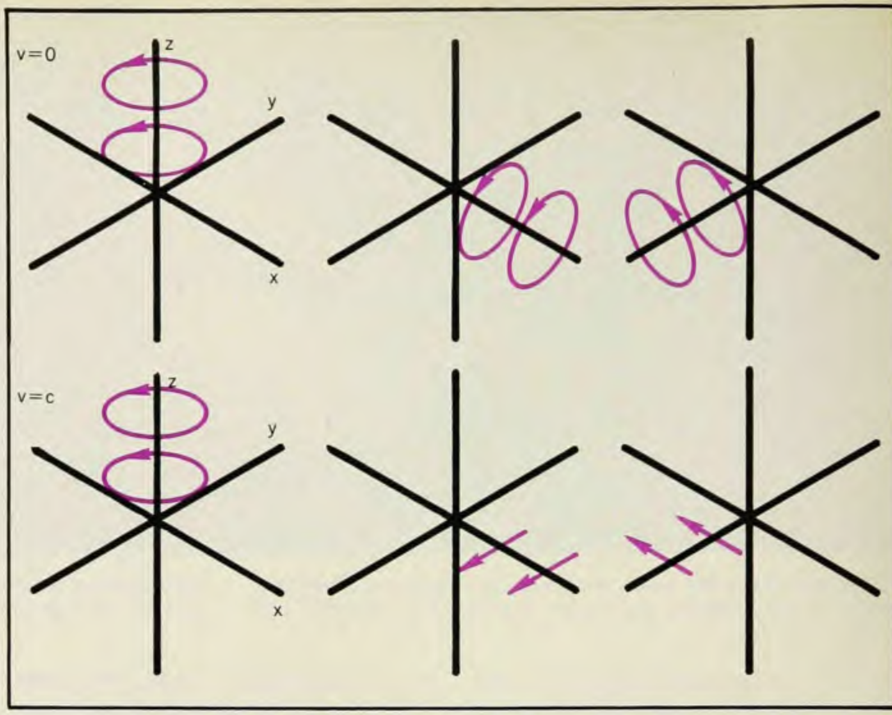
Having argued that electrodynamics and optics can be understood in terms of the photon as just another particle, but one that happens to have the particular quantum numbers of mass zero and spin one, I will now contradict myself and point out why that is not the most fruitful approach. Let us turn the argument around and say that gauge invariance is a fundamental symmetry of nature. From gauge invariance, now taken as an *a priori* assumption, we can derive the fact that there should exist a particle, the photon, with mass zero and spin one, and furthermore derive Maxwell's equations as well as all the other properties of electrodynamics.

Which way should we go, then? From the known quantum numbers to gauge invariance, or from gauge invariance to the known quantum numbers? Obviously questions like these do not make any sense in terms of predictions of experimental data, but they make a lot of sense in terms of the direction theoretical research is likely to take.

I would like to argue that it makes sense to talk about gauge invariance as a fundamental symmetry of nature from which follows the existence of the photon and therefore also light and Maxwell's equations. The justification for this point of view is the fruitfulness of the idea—its fruitfulness, to be specific, in unifying electromagnetism with other areas of physics. To understand this, let us recall the way in which elementary-particle physicists these days categorize the various kinds of interactions enjoyed by these particles.

The properties of the four types of interactions are listed in Table 2. We have already discussed the gravitational and electromagnetic interactions, which are the only ones felt in everyday life. They have ranges that, as far as we know, are infinite. They are characterized by strengths that are quite different, the gravitational being very much weaker than the electromagnetic. Gravitation, however, makes up for its weakness by affecting everything, whereas electromagnetism only affects charged particles. I have mentioned above that we now believe gravitational forces to be transmitted by the exchange of gravitons, and we know that electromagnetic forces are transmitted by the exchange of photons.

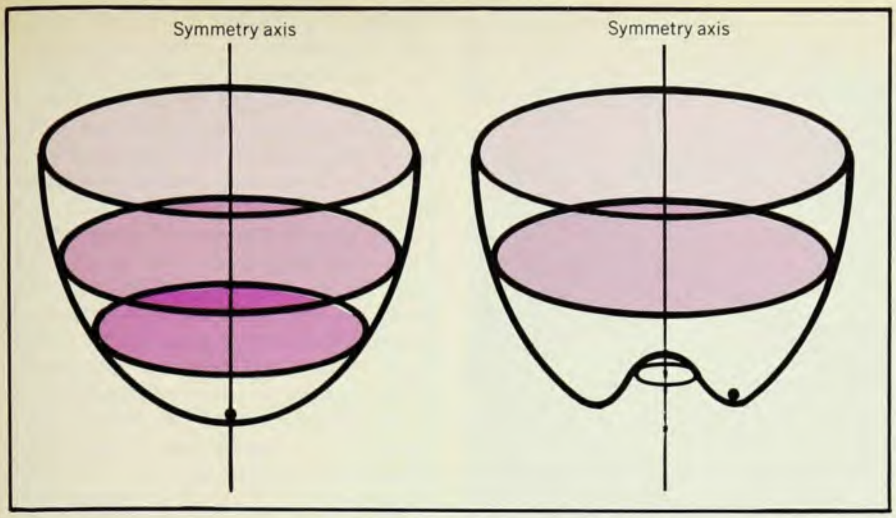
The strong interactions, which are supposedly responsible for holding the nucleus together, are much less familiar. They are two orders of magnitude stronger than the electromagnetic, and



Why the photon has only two polarization states. Of the three circles (top) representing rotation components of a particle at rest, two (the ones normal to the motion) are squashed into lines for the massless photon, which travels at the speed of light. Figure 2

Table 1. Some elementary particles and their properties

Particle	Charge (e)	Spin ($\hbar$)	Mass (MeV/c^2)	Lifetime (sec)
Photon	0	1	0	∞
Graviton (?)	0	2	0	∞
Neutrino	0	1/2	0	∞
Electron	±1	1/2	0.511	∞
Muon	±1	1/2	106.66	2.199×10^{-5}
π meson	±1	0	139.576	2.602×10^{-8}
	0	0	134.972	0.84×10^{-16}
K meson	1	0	493.84	1.237×10^{-8}
	0	0	497.79	0.862×10^{-10}
η meson	0	0	548.8	2.50×10^{-17}(?)
Proton	1	1/2	938.259	∞
Neutron	0	1/2	939.553	935
Λ hyperon	0	1/2	1115.59	2.521×10^{-9}
—	—	—	—	—



Spontaneous symmetry breaking is here illustrated by a mechanical analogy. The stable equilibrium condition of the ball in the bowl on the left is on the axis of cylindrical symmetry; when the bottom is indented (right), a circular trough becomes stable instead. In the analogy the position of the ball corresponds to the vacuum expectation value of a scalar field, the oscillation frequencies to scalar particle masses and the newly created rolling mode to a Goldstone boson of zero mass—the π meson can be regarded as one of these. Figure 3

Table 2. The four interactions

	Gravitational	Electro-magnetic	Strong	Weak
Range	∞	∞	10^{-13}–10^{-14} cm	$\ll 10^{-14}$ cm
Examples	Astronomical forces	Atomic forces	Nuclear forces	Nuclear beta decay
Strength ($\hbar = c = m_p = 1$)	$G_{\mathrm{Newton}} = 5.9 \times 10^{-39}$	$e^2 = 1/137$	$g^2 \approx 1$	$G_{\mathrm{Fermi}} = 1.02 \times 10^{-5}$
Particles acted upon	Everything	Charged particles	Hadrons	Hadrons and leptons
Particles exchanged	Gravitons	Photons	Hadrons	?

Table 3. Gauge groups

Group	Vector fields	Physical application
O(2)	A_μ	Electromagnetism
O(3)	A_μ, $W_\mu^\pm$	Yang–Mills theory of strong interactions
O(3) $\otimes$ O(2)	A_μ, $W_\mu^\pm$, Z_μ	1967 model of weak and electromagnetic interactions

they act on a class of particles called hadrons, which includes the proton, the neutron and various mesons and hyperons. The strong interactions are believed to be transmitted by the exchange of hadrons.

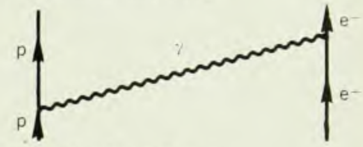
The least familiar of all are of course the forces that have the shortest range; these are the weak interactions. While the strong interactions have ranges of the order of 10^{-13}–10^{-14} cm and so are only important inside the nucleus, the weak interactions have even a very much shorter range and, so far as we know, are not responsible for holding anything together. They are, however, responsible for nuclear beta decay, including some of the reactions that are responsible for producing the heat of the Sun. These forces are very much weaker than the electromagnetic interactions. Although we know that they act upon hadrons and leptons (electrons, muons and neutrinos), we do not quite know what particles are exchanged in giving rise to these weak interactions.

There is a rule that the range of the force produced is inversely proportional to the mass of the exchanged particle:

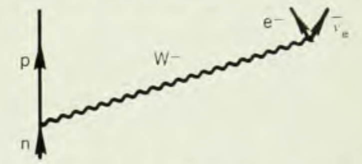
$$\text{range of force} \approx (\hbar/c)/\text{mass of exchanged particle}$$

That is why we believe that there is a massless graviton in addition to the massless photon. It also suggests that a particle must be very massive if its exchange is to be responsible for the weak interactions.

It is, in fact, an old idea that, just as Coulomb scattering, which is an electromagnetic process, proceeds by the exchange of photons, so beta decay proceeds by the exchange of a heavy kind of "photon." We can deduce the properties of the W particle, the supposed intermediate particle of beta decay, from the analogy between the electromagnetic and the weak interaction. Let us compare the Feynman diagram for Coulomb scattering



with that for beta decay



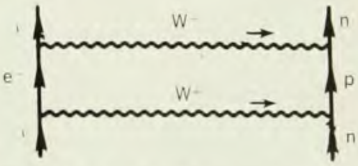
By analyzing the polarization states of the various particles involved in beta decay, we conclude that the spin of the W particle, if it exists, has to be one, the same as that of the photon. Integer spin makes it a boson and, because it has to be described in terms of the vector field, like the photon, it is called the intermediate vector boson. The W particle is, however, very different from the photon in that it has plus or minus one unit of charge. Furthermore, from the known strength of the interaction, we deduce that its mass is roughly 50 GeV.

These ideas go back to Enrico Fermi, with contributions since then by Hideki Yukawa, Julian Schwinger, Sheldon Glashow, Sydney Bludman, Abdus Salam and John Ward, and others.

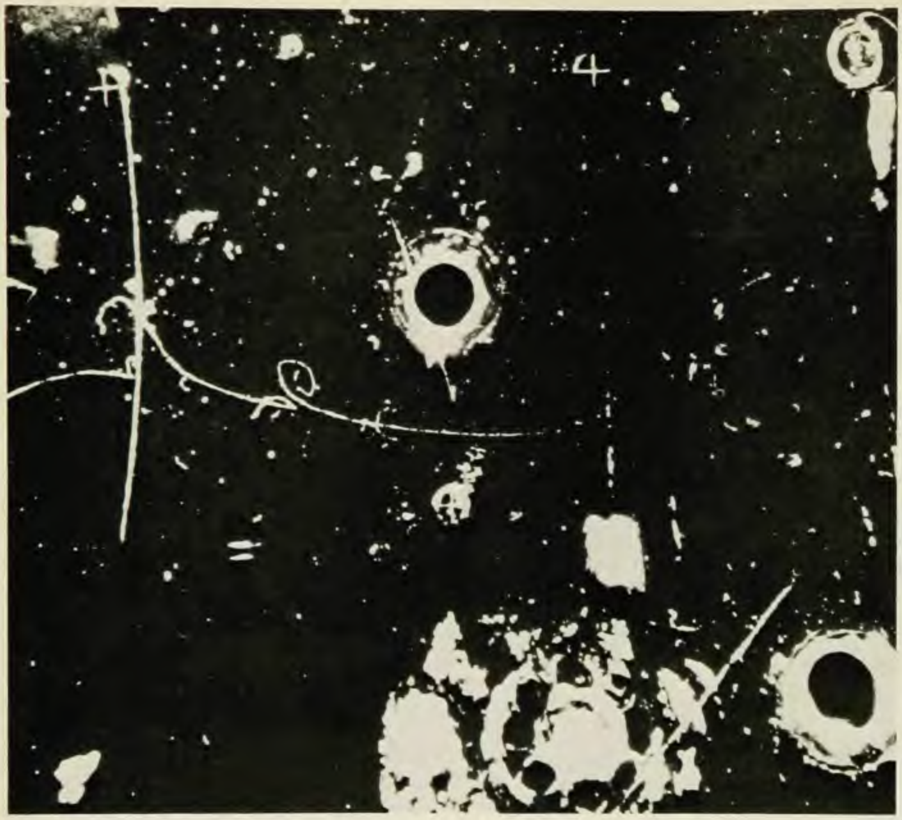
Generalized gauge invariance

There are difficulties, however, with these ideas. One of the problems is that we are putting two radically different particles together into an analogous relationship: the photon, with zero mass, and the W particle, which has a mass of about 50 proton masses, much heavier than anything we have ever seen.

An equally important, although less obvious, problem is the appearance of infinities. Imagine the scattering, say, of a neutrino on a neutron. This is a process that normally would be expected to go by the exchange of a *pair* of W particles, since one of them alone can not do it. If you analyze this Feynman diagram:



you find that it is infinite and that there



Detail of a photograph taken at the Gargamelle bubble chamber at CERN. A beam of muon antineutrinos enters the chamber from the right. One strikes an electron in an atom of the fluid, causing it to recoil to the left, spiralling clockwise in the chamber's magnetic field before it finally comes to rest. The tracks immediately to the left of the recoil-electron track are produced by an electron–positron pair that was created in the electric field of an atom by a photon emitted by the decelerating recoil electron. Figure 4

is no way of absorbing the infinity into a renormalization of coupling constants. When you further analyze this infinity, you find that it arises precisely from the exchange of W particles with helicity zero. For photons this diagram is not infinite because photons, as we saw, have no states of helicity zero.

In 1967 I made a proposal, which was subsequently also made by Salam, to derive the photon and the W as siblings from a general gauge-invariance principle—one that would be a generalization of the gauge-invariance principles I have described above in connection with electrodynamics. In this theory the mass of the W particle would arise from what is known in the theoretical trade as spontaneous symmetry breaking. This idea is that, on some fundamental level—on the level of the laws of nature, although not on the level of the predicted matrix elements of the scattering amplitudes—the photon and the W can be seen as members of the *same family* of vector particles.

A simple illustration of spontaneous symmetry breaking is provided by the two bowls of figure 3. In the analogy, the projection of the position of the ball onto the horizontal plane corresponds to the vacuum expectation value of a complex scalar field ϕ, and the oscillation frequencies to masses of a pair of scalar particles. When the bottom of the bowl is pushed in to form a dimple, the equilibrium point spreads out into a circle—the former symmetry is broken. The resulting new rolling mode of the ball about the axis of cylindrical symmetry represents, in this analogy, a new particle, called a Goldstone boson of zero mass. The pi meson may be regarded, at least approximately, as such a Goldstone boson.

Such an extended particle family would be required by a gauge-invariance principle broader than the gauge-invariance principle of electrodynamics.

The kind of generalized gauge invariance that I am referring to is this: Imagine a group, not just a simple group of phase transformations such as that of equation 1, but a group that has to be represented by two noncommuting matrices, M and $\mathfrak{M}$. The various charged fields transform according to the matrix transformations

$$\psi_n(x) \longrightarrow \sum_m M_{nm}[\phi(x)]\psi_m(x)$$

$$A_{\alpha\mu}(x) \longrightarrow$$
$$\sum_\beta \mathfrak{M}_{\alpha\beta}[\phi(x)][A_{\beta\mu}(x) + \partial\phi_\beta(x)/\partial x^\mu]$$

The $\phi_\alpha(x)$ are a set of arbitrary functions of position.

There would have to be one vector field for every one of the parameters that characterizes the group transformation, which would themselves form a nontrivial family. More than one member of this family would have to undergo a matrix transformation as well as the familiar gauge transformation.

This kind of generalized gauge invariance was first brought into physics by Chen Ning Yang and Robert Mills in the 1950's, and has been kicking around as a mathematical possibility for many years. The suggestion that was made in 1967 was in fact to see the photon as part of the family of vector fields required by a fundamental and *exact* gauge invariance of nature. The kind of gauge invariances that appear are shown in Table 3. There is, first, the gauge group 0(2) which, on the Argand diagram, is just a group of phase transformations. It is the same as the group of rotations in two dimensions, which would give rise only to a single vector field (there is only one way you can rotate in two dimensions) and, as I pointed out above, we identify this vector field with the photon.

The original suggestion of Yang and Mills really had to do with the strong interactions and isotopic spin. The idea was that the photon would be accompanied by a pair of charged vector fields, and the gauge group would be the group of rotations in three dimensions, 0(3).

New kinds of interactions

The particular proposal made in 1967, the simplest one possible, is that the fundamental gauge group is the direct product, 0(3) ⊗ 0(2), of 0(3) and 0(2). As a result, instead of three vector fields, there were four. This would imply that, in addition to the heavy charged intermediate vector bosons that would be responsible for the observed weak interactions such as beta decay and the nuclear reactions in the Sun, there would be an additional neutral vector meson, similar to the photon in having no charge, but unlike it in being heavy like the W's.

This 1967 proposal went nowhere for a long time, because, although it was suggested that it would eliminate the infinities from the theory of weak interactions, for a long time no one was clever enough to prove it. This was done in 1971 by Gerhard 't Hooft, then a graduate student at Utrecht, and his proof was later improved by other people including Benjamin Lee. As soon as it became clear that this theory was in fact a solution to the problem of the infinities in the weak interaction, one that had been with us since the mid-1930's, physicists began to put a tremendous amount of work into it, both in examining its applications and in looking for experimental tests.


OBSERVATION OF NEUTRINO-LIKE INTERACTIONS WITHOUT MUON OR ELECTRON IN THE GARGAMELLE NEUTRINO EXPERIMENT

Fifty-five authors had a hand in writing this three-page paper from Gargamelle, in which neutral-current events were reported. Is this a record? Figure 5

There are now a number of experimental reasons for believing that this theory is in some sense correct. One of them is that theories of this type—and there is more than one theory of this type—predict the remarkable fact that charge must come in quantized units.

Furthermore, these theories tend to predict that the weak interactions should have some effects comparable in strength to the effects of electromagnetism, but only of certain very limited types. For example, there is strong reason to believe that the mass difference between the neutron and the proton, which has usually been thought of as entirely electromagnetic (arising from the fact that the proton has a charge and the neutron does not) is in fact partly electromagnetic but mostly weak.

But there is a prediction, more accessible to direct test, of this kind of theory —there should be new kinds of weak interactions produced by the exchange of the heavy neutral intermediate vector particles; among these are the scattering of a muon type of neutrino by an electron, which has now apparently been observed in two events at CERN. A bubble-chamber photograph of one of these "neutral-current" events is shown in figure 4. Two events seem like a very small number, but the background is very low in this kind of experiment.

On the other hand, there are now hundreds of events in which a neutrino is seen to produce hadrons in a collision with a nucleon; in these, no exchange of charge takes place between the neutrino and the hadrons. Even though hundreds of events have been observed, both at CERN and at the National Accelerator Laboratory, the background in this case is a serious problem—so that this is also not conclusively settled. For its sociological interest, figure 5 shows what an experimental paper in high-energy physics looks like these days—there are 55 authors!

[Another recent encouraging development is the discovery of new long-lived vector particles at SLAC and Brookhaven. These new particles may be bound states of a new kind of "charmed" quark and the corresponding antiquark; such a charmed quark is needed in unified theories of weak and electromagnetic interactions to suppress certain unobserved neutral-current process.]

There has also recently been a breakthrough in the work of Daniel Freedman and James Wilson and others on the question of how supernovas manage to explode. The key idea that is involved is that the neutral currents produce a coherent interaction between the neutrinos and the iron nuclei in the outer core of a supermassive star. For the first time, this interaction provides a computer model that is actually capable of blowing the star apart and producing the observed phenomenon of a supernova, such as the one that occurred in the year 1054, the remains of which is the Crab nebula shown in figure 1.

The symmetry of nature

How then do we answer the question, What is light? The answer in which I now have the greatest faith is: The photon is the most visible member of the family of elementary particles required by a generalized gauge group that mediates the electromagnetic, weak and perhaps also the strong interactions. As far as I have seen, no one has claimed to be able to include gravitation in this scheme.

If these theoretical ideas, and the experiments that are going on, pan out, we will begin to understand in a fundamental way what light is, and that the photon (as well as other particles that are far less familiar because we live on a much longer time scale than they do) forms the manifestation of a symmetry principle of nature that describes the interactions of matter. This principle is about as fundamental as anything we know about the world.

Bibliography

- S. Weinberg, in *Lectures on Particles and Field Theory* (S. Deser, K. W. Ford, eds.), Prentice-Hall, Englewood Cliffs, N.J. (1965), vol. 2, page 405.
- E. S. Abers, B. W. Lee, Phys. Reports **9**, 1 (1973).
- J. Bernstein, Rev. Mod. Phys. **46**, 7 (1974).
- S. Weinberg, Rev. Mod. Phys. **46**, 255 (1974).
- S. Weinberg, Scientific American, July 1974, page 50. □