# A question of mass

Jeremy Bernstein[a)]

(Received 7 May 2010; accepted 17 August 2010)

We present a pedagogical discussion of spontaneous symmetry breaking, the Goldstone theorem, and the Higgs mechanism. If the Higgs boson is found, it might provide an explanation of the origin of mass. 
[DOI: 10.1119/1.3487939]

"The quantity of any matter is the measure of it by its density and volume conjointly…. This quantity is what I shall understand by the term *mass* or *body* in the discussions to follow. It is ascertainable from the weight of the body in question. For I have found by pendulum experiments of high precision, that the mass of a body is proportional to its weight; as will hereafter be shown." Isaac Newton[1]

When I took freshman physics as a sophomore at Harvard in 1948, this definition of mass was still used in our textbook. As it happens, the previous year I had taken a philosophically oriented course in modern physics given by the philosopher-physicist Philipp Frank. He introduced us to the work of his fellow Austrian philosopher-physicist Ernst Mach. Therefore I knew of Mach's devastating critique in his book, *The Science of Mechanics*,[1] and how it had influenced Einstein. Newton's definition of mass is circular. What is density? How does it is apply to the photon, which has no mass?

When I began writing my Ph.D. thesis in the early 1950s, I would have described myself as an "elementary particle" theorist rather than as a nuclear theorist. Elementary particles were considered to not consist of anything else, whereas the atomic nucleus consists of neutrons and protons, which were taken to be elementary particles. There were not that many elementary particles known at the time. In addition to the neutron and the proton, there was the photon, the electron, and the neutrinos. The positron was known, and most physicists, Feynman being a notable exception, believed that the antiproton would be found once accelerators were sufficiently energetic. A few years earlier, a "heavy electron" had been discovered that had some of the properties of the electron, including its weak and electromagnetic interactions, except that it was about two hundred times more massive and unstable. For various reasons that no longer make any sense, it was first called the mu meson and then eventually the muon. It seemed to serve no purpose and when I. I. Rabi heard of it, he asked, "Who ordered that?"

Rabi's pique was understandable. In the 1930s, a theory of the nuclear force had been proposed. It had to account for satisfying two conditions. First, the nuclear force was very short ranged and acted only when the neutrons and protons were practically on top of each other. Second, it had to be much stronger than the electrical force; otherwise, the positively charged protons, which repel each other, would tear the nucleus apart. As it is, heavy nuclei with many protons tend to fission spontaneously. Both of these conditions could be satisfied if a fairly massive particle was exchanged between the neutrons and protons and among themselves. The strength of this interaction was postulated to be large compared to the electrostatic force. It was also shown that the range $r$ of this force was related to the mass $m$ of the particle being exchanged. From the uncertainty principle for energy and time, with the energy uncertainty equal to $mc^2$, we have $mc^2 \sim \hbar c/r$, and thus its mass was predicted to be about 400 times larger than that of the electron. This mass, unlike the mass of the muon, seemed to have a connection to the dynamics. The muon had something like the correct mass, but it only interacted electrically and weakly, and thus it was the wrong particle. The right particle was called the pi-meson or the pion. Why did it not show up in cosmic rays rather than the muon? The answer turned out to be very simple. The pion, when it was not absorbed in the atmosphere, decayed into a muon and neutrinos. When accelerators became sufficiently powerful, they produced pions in droves. At the time I was writing my thesis, pion physics was flourishing. But then, the roof fell in.

Particles that no one anticipated also began to show up in droves in cosmic rays. They became known as "strange particles" because they were. There was the K meson, which came in a charged and neutral variety, and hyperons, which had masses greater than the proton or neutron. The latter category includes a lambda particle, which is neutral, a sigma particle with charges plus, minus, and zero, and a xi particle with charges zero and minus. These were the lowest mass particles, which were repeated in higher mass replicas. In short, it was a particle zoo. It took strong nerves to see any pattern, a pattern that might reflect an underlying symmetry.

At the time, there was a clear idea of how such a symmetry might appear. The neutron and proton were prime examples. They had many properties in common including their spins, which were identical, and their masses, which did not differ by much. The neutron was a bit heavier and decayed into a proton, an electron, and an antineutrino. Suppose we imagined a world in which electromagnetism was switched off. In this world, the neutron and proton would have the same mass and would collapse into a doublet. The three pi mesons, plus, minus, and zero charge, would collapse into a triplet. A symmetry would emerge, which was called "isotopic spin," which is invariance under the group SU(2). The predictions were reasonable, and hence isotopic spin was a useful approximate symmetry. Maybe something analogous could be found for the strange particles.

The trouble was that the mass differences were too great. Although the mass difference of the neutron and proton was only a fraction of a percent of the mass of either particle, the K mesons had nearly four times the mass of the pions. It took

an act of faith to see how these objects fitted into some kind of symmetric structure. But Murray Gell-Mann took the leap. He proposed a symmetry that was a generalization of isotopic spin, SU(3), and suggested that if the various mass differences were neglected, the known particles could be organized in multiplet structures. For example, the known scalar mesons, including a newly discovered particle that was called the eta, fitted into an octet. The known hyperons fitted into a tenfold decuplet, but there was one missing. It was given the name $\Omega^-$ and its properties were predicted. When it showed up with these properties, the lingering doubts about this scheme vanished and Gell-Mann was awarded a well-deserved Nobel Prize. It was a textbook example of a symmetry and its breaking.

This kind of symmetry breaking is nearly as old as the quantum theory itself. Eugene Wigner and Herman Weyl, for example, studied the role of group theory in quantum mechanics. The idea was that the description of a quantum mechanical system could be split into two contributions—a Hamiltonian that exhibits the symmetries of the group plus a second Hamiltonian that did not. If the later is "small," then some aspects of the original symmetry would still be apparent.

As an example, consider the group of rotations in three-dimensional space called SO(3). These rotations are generated by the orbital angular momentum. Suppose one part of the Hamiltonian contains only a central force. This part is invariant under rotations, which means that the angular momentum operators commute with this part of the Hamiltonian. The eigenstates are both eigenstates of the energy and the angular momentum. If, for example, the nuclear force that binds the neutron and proton together is represented by a central force, then the ground state of the deuteron would be an S-state. But it isn't. It has a small percentage of the D state, which manifests itself in the fact that the deuteron has a quadrupole moment. The rotational symmetry is broken in this case by adding a tensor force. The total angular momentum, which includes the spin, is conserved, but the purely orbital part is not. Nonetheless, it is still useful to expand the wave functions in eigenfunctions of the angular momentum.

In the isotopic spin example, the neutron-proton system in the absence of electromagnetism shows symmetries under the group of special unitary transformations SU(2). Once electromagnetism is included, the symmetry is broken, but nonetheless, there are still some useful manifestations. Likewise, the elementary particles in the absence of symmetry breaking are invariant under the special unitary group SU(3). If this symmetry is broken, it is still possible to derive relations among the masses, but their origin was still unexplained. However, in the early 1960s, Yoichiro Nambu and others showed that a second kind of symmetry breaking was possible in quantum mechanics, which was called spontaneous symmetry breaking. To see what it means, we consider an example that has nothing to do with quantum mechanics.

Consider the equation

$$d^2f(x)/dx^2 = x^2 + cx. \tag{1}$$

If $c=0$, Eq. (1) is symmetric under $x$ inversion: $x \to -x$. We drop the integration constants and obtain

$$f(x) = 1/12x^4 + c/6x^3. \tag{2}$$

The solution with $c \neq 0$ is not $x$ inversion symmetric. This, lack of symmetry is in the spirit of Wigner-Weyl symmetry breaking. But consider

$$d^2f(x)/dx^2 = x^2. \tag{3}$$

Equation (3) is $x$-inversion symmetric. The solution is

$$f(x) = 1/12x^4 + Bx + C. \tag{4}$$

Unless $B$ is zero, the solution does not have the same symmetry as Eq. (3). The symmetry has been "spontaneously broken" by the choice of solution, determined by the initial conditions.

To see how spontaneous symmetry breaking works in quantum mechanics and to understand the consequences, we consider an example, the self-interactions of a complex scalar field, $\Phi(x,t)=\Phi_1(x,t)+i\Phi_2(x,t)$, where $\Phi_{1,2}$ are real fields. This field describes a charged spinless particle. It is the simplest example that I know, and it is the one that was first considered historically. See, for example, Ref. 2. It will lead us to the Higgs mechanism.

We begin by exhibiting the Lagrangian of a free complex scalar classical field, which corresponds to a particle with a mass $m$. To make the notation more compact, I will employ the usual convention of setting $c=1$. Thus

$$\begin{aligned} L &= \partial_\mu\Phi(x)^\dagger\partial_\mu\Phi(x) - m^2\Phi(x)^\dagger\Phi(x) \\ &= (\partial\Phi/\partial t)^\dagger(\partial\Phi/\partial t) - (\nabla\Phi)^\dagger(\nabla\Phi) - m^2\Phi^\dagger\Phi, \end{aligned} \tag{5a}$$

and the corresponding Hamiltonian is

$$H = (\partial\Phi/\partial t)^\dagger\,\partial\Phi/\partial t + (\nabla\Phi)^\dagger \bullet\ \nabla\Phi) + m^2\Phi^\dagger\Phi. \tag{5b}$$

We shall be interested in minimizing the energy. The kinetic terms are always positive definite, and therefore to minimize the energy associated with $H$, we must take the fields to be constants in spacetime.

The Lagrangian in Eq. (5a) yields the field equation

$$(\partial_t^2 - \nabla^2 - m^2)\Phi(x) = 0, \tag{6}$$

where I have simplified the notation by using $x$ for $\{x,t\}$. The solutions of the field equations for individual particle and antiparticle states with momentum $p$ have energies $\sqrt{p^2+m^2}$, which establishes the interpretation of $m$ as a particle mass.

The Lagrangian is invariant under the global gauge transformation $\Phi \to \exp(i\Lambda)\Phi$, where $\Lambda$ is a real number. It is not invariant under the spacetime dependent or local gauge transformation $\Phi \to \exp(i\Lambda(x))\Phi$. Under an infinitesimal transformation, a term is added to the Lagrangian of the form, with $\Lambda_{,\mu}$ standing for the derivative with respect to the $\mu$th coordinate

$$\delta L = i\Lambda_{,\mu}(\partial_\mu\phi^\dagger\phi - \phi^\dagger\partial_\mu\phi) \equiv \Lambda_{,\mu}J_\mu. \tag{7}$$

If we think of $\Lambda$ as a dynamical variable, then we can write down the Euler–Lagrange equation for $\Lambda$ as

$$\partial_\mu\delta L/\delta\Lambda_{,\mu} = \delta L/\delta\Lambda. \tag{8}$$

The term on the left is the divergence of the current generated by the local gauge transformation, and the term on the

right vanishes, and hence the current generated this way is conserved. This argument is a variant of what is known as a Noether theorem.

From the Noether theorem, we have

$$\partial_\mu J_\mu = \nabla \cdot J(x,t) + \partial/\partial t J_0(x,t) = 0, \tag{9a}$$

$$0 = \int d^3x \partial_\mu J_\mu. \tag{9b}$$

If nothing strange happens at the boundary in space, Eq. (9b) reduces to

$$0 = \partial/\partial t \int d^3x J(x,t)_0 \equiv \dot{Q}, \tag{10}$$

where $Q$ is the spatial integral of the charge density $J_o(x,t)$. We now want to quantize this theory. Thus we write

$$\Phi(x) = \int d^3k (b^\dagger{}_k \Phi_k{}^{(-)}(x) + a_k \Phi_k{}^{(+)}(x)). \tag{11}$$

Here, $\Phi$ in the integral can be taken as a shorthand for the Fourier transform coefficients, say, $\exp(ikx)$, with some suitable normalization. The $a$ and $b$ coefficients are annihilation operators, and their conjugates are creation operators. The only nonvanishing commutators of the $a$ and $b$ operators are

$$[a_k, a_{k'}{}^\dagger] = \delta^3(k-k') = [b_k, b_{k'}{}^\dagger]. \tag{12}$$

The momentum $\pi(x)$ conjugate to $\Phi$ is given by

$$\pi(x) = \hbar \partial_t(\Phi^\dagger(x))$$
$$= -i\hbar \int d^3k \omega_k (b_k \Phi_k{}^{(-)*}(x) - a_k{}^\dagger \Phi_k{}^{(+)*}(x)). \tag{13}$$

Equation (13) yields the commutation relation

$$[\Phi(x), \Pi(y)] = i\hbar \delta(x-y). \tag{14}$$

The charge can be written in terms of the $a$ and $b$ operators as

$$Q = \int d^3p (a_p{}^\dagger a_p - b_p{}^\dagger b_p). \tag{15}$$

From Eq. (15), it follows that

$$[Q, \Phi] = \Phi \tag{16}$$

so that

$$\exp(i\Lambda Q)\Phi \exp(-i\Lambda Q) = \exp(i\Lambda)\Phi. \tag{17}$$

Thus the charge generates the global gauge transformation, that is, $\Phi$ transforms under unitary transformations generated by the operator $\exp(i\Lambda Q)$.

The Hamiltonian in normal ordered form with $a^\dagger$ and $b^\dagger$ to the left of $a$ and $b$ is given by

$$H = \int d^3k \omega_k (a_k{}^\dagger a_k + b_k{}^\dagger b_k). \tag{18}$$

We have dropped an additive constant. Additive constants to the Hamiltonian do not change the physics because we are concerned with energy differences and not with absolute energies. (In contrast, an additive constant to the Lagrangian changes the action, and thus the transition amplitudes that are calculated with path integrals. We must make sure that additive constants do not change the physics by eliminating any constants.)

The vacuum state $|0\rangle$ is that state for which the energy is minimum. If it has the property that $\langle 0|H|0\rangle = 0$, it follows that from equations such as $\langle 0|a_k^\dagger a_k|0\rangle = 0$ that

$$a_k|0\rangle = b_k|0\rangle = 0, \tag{19}$$

and thus both the Hamiltonian and the charge operators acting on this state are zero. We also have the obvious property

$$\langle 0|\Phi|0\rangle = 0. \tag{20}$$

So far the particle mass has been specified without any explanation for its origin. We now want to introduce mass generation through spontaneous symmetry breaking. We introduce a new Lagrangian

$$L = 1/2 \partial_\mu \Phi(x)^\dagger \partial_\mu \Phi(x) + m^2/2 \Phi(x)^\dagger \Phi(x)$$
$$- \lambda/4 (\Phi^\dagger(x)\Phi(x))^2. \tag{21}$$

Several properties of this Lagrangian are evident. First, the term proportional to $m^2$ is not a mass term. Compare the sign to the sign in the mass term in Eq. (5a). Instead, it is a self-interaction term. Second, the Lagrangian is invariant under global gauge transformations but not local ones. There is a conserved current as before and a conserved charge. But does this charge annihilate the vacuum, that is, $Q|0\rangle = 0$ as before, and if not what does this mean? Here we run into the question of what is the vacuum.

We recall the equation

$$[Q, \Phi] = \Phi, \tag{22}$$

which also holds here. Equation (22) implies that if $Q$ does not annihilate the vacuum, then it must be that

$$\langle 0|\phi|0\rangle \neq 0. \tag{23}$$

Equation (23) means that $\Phi$ cannot have a particle interpretation because we cannot build up the single particle states by the creation operators acting on the zero particle vacuum state.

We recall from the discussion following Eq. (5a) that the energy is minimized for constant fields, and it is therefore determined by minimizing the potential. Let us consider the classical potential

$$V = -m^2/2(\Phi^\dagger\Phi) + \lambda/4(\Phi^\dagger\Phi)^2. \tag{24}$$

Clearly one extremal is when $\Phi = 0$. Quantum mechanically we want to replace this condition by the condition that the vacuum expectation value of the potential is a minimum. We shall see that in this case, there is no unique answer.

Let us warm up with a simpler case, which will illustrate the issues. We consider a real field $\Phi$ and the potential

$$V = -m^2/2\Phi^2 + \lambda/4\Phi^4. \tag{25}$$

The potential and $\Phi$ are all functions of the spacetime point $x$. At the minima of the vacuum expectation value of the energy, they are constants, and hence we can find a condition

on $\Phi$ that minimizes the potential for all $x$. We take the derivative with respect to $\Phi$ and set it equal to zero. Thus

$$\Phi(m^2 - \lambda\Phi^2) = 0. \tag{26}$$

Equation (26) has three solutions,

$$\Phi = 0, \pm \sqrt{m^2/\lambda}. \tag{27}$$

Equation (27) corresponds to the values of the potential at 0, which is a local maximum and $\pm 1/4\ m^2/\lambda$, two distinct minima with the same energy. If we pick the one with the positive minimum for $\Phi$, then for this vacuum

$$\langle 0|\Phi|0\rangle = \sqrt{m^2/\lambda} \equiv v. \tag{28}$$

Equation (28) shows that $\Phi$ does not have the usual particle interpretation and suggests that we introduce a new field $\eta$ to describe the fluctuations of $\Phi$ away from its constant vacuum value $v$. We have

$$\eta = \Phi - v. \tag{29}$$

In terms of $\eta$, the Lagrangian becomes

$$L = 1/2\partial_\mu\eta\partial_\mu\eta - m_\eta{}^2\eta^2 - \lambda v\,\eta^3 - 1/4\lambda\,\eta^4 + m^4/4\lambda, \tag{30}$$

where

$$m_\eta = \sqrt{2m^2}. \tag{31}$$

This choice of vacuum has produced an $\eta$, a particle with a nonzero mass and some peculiar self interactions. But note that the $\Phi \rightarrow -\Phi$ symmetry of the original Lagrangian in Eq. (21) has been broken spontaneously. There is no trace of it in the transformed Lagrangian. The last term, which is a constant, also deserves further comment. If we were considering a Hamiltonian, we could add a constant term with no misgivings. But as I have mentioned, the Lagrangian is different. From it, we define the action $\int_t^{t'} L dt$. If we add a constant to the Lagrangian, it adds a term proportional to the time difference in the action. We had better eliminate this term if we want a sensible theory.

With this example in mind, we now return to the complex fields with the continuous global gauge transformation invariance. As we shall see, this invariance brings in something new. The way to deal with this case is to write

$$\Phi = 1/\sqrt{2}(\Phi_1 + i\Phi_2), \tag{32}$$

where $\Phi_{1,2}$ are real fields. In terms of these fields, the Lagrangian becomes

$$L = 1/2(\partial_\mu\Phi_1)^2 + 1/2(\partial_\mu\Phi_2)^2 + 1/2m^2(\Phi_1^2 + \Phi_2^2) - 1/4\lambda(\Phi_1^2 + \Phi_2^2)^2. \tag{33}$$

The minima are given by the condition that

$$\Phi_1^2 + \Phi_2^2 = m^2/\lambda \equiv v^2. \tag{34}$$

The phase is undetermined. We choose the phase so that at the minimum,

$$\Phi_1 = \sqrt{m^2/\lambda}, \quad \Phi_2 = 0. \tag{35}$$

We can then displace $\Phi_1$ by its vacuum expectation value in the vacuum defined by this choice of phase and thus write

$$\Phi(x) = (1/\sqrt{2})[v + \eta(x) + i\xi(x)]. \tag{36}$$

We can rewrite the Lagrangian in terms of these fields. There will be self-interaction terms of $\{\eta, \xi\}$ as well as interactions between them and the additive constant. What interests us is the "kinetic" term $L_K$,

$$L_k = 1/2(\partial_\mu\xi)^2 + 1/2(\partial_\mu\eta)^2 - m^2\eta^2, \tag{37}$$

which shows that the new $\xi$ field is massless and the $\eta$ field has mass $m$.

Let us review what we have done. We began with a Lagrangian for a complex field of zero mass, which was globally gauge invariant. We broke this gauge invariance spontaneously and found two interacting real scalar fields. One of these fields has mass zero, and the other has acquired a mass. Is this result some freakish artifact of this Lagrangian, or are we in the presence of a more general phenomenon? The answer is the latter. We have found a realization of what is known as the Goldstone theorem.

I will not try to give a detailed proof of this theorem here but only state what it is. There are fine points that I will discuss shortly. Suppose you have a theory with a certain number of conserved currents, and these currents give rise to conserved charges that generate some set of gauge transformations. If one of these charges has a nonvanishing expectation value so that the gauge symmetry is broken spontaneously, then necessarily it will give rise to a mass zero, spin zero particle—the $\xi$ in the example we have discussed. On its face, this result would appear to rule out theories of this kind in elementary particle physics because there are no such particles. However, there is a loophole, and through it we will drive a truck. The loop hole is Lorentz invariance.

Needless to say, we want all our theories to be Lorentz invariant, but they need not be "manifestly" Lorentz invariant. A case in point is electrodynamics. This theory is certainly Lorentz invariant. When Einstein had to choose between Newtonian mechanics and electromagnetism, he chose the latter precisely because it was relativistic. But electromagnetism is not manifestly Lorentz invariant in the following sense. The photon field $A_\mu$ is not well-defined. The theory is invariant under gauge transformations of the form $A_\mu \rightarrow A_\mu + \partial_\mu\Lambda$, where $\Lambda$ is a function of the spacetime point $x$. This invariance precludes terms such as $A_\mu A_\mu$ in the Lagrangian, and thus the photon has no mass.

To define the theory, we must select a gauge. Two popular gauges are the Lorenz gauge,[3] with $\partial_\mu A^\mu = 0$, and the Coulomb gauge with $\nabla \cdot A = 0$. The Lorenz gauge condition is manifestly Lorentz invariant, and the Coulomb gauge is not. You can use either gauge to carry out calculations. You will get the same answers for any physical quantity, and these answers will be Lorentz covariant.

The proof of the Goldstone theorem that most clearly makes use of the manifest Lorentz covariance is due to Walter Gilbert.[4] Gilbert has an interesting history. He got his Ph.D. in physics from Abdus Salam and then switched into biology. In 1980, he won the Nobel Prize for chemistry. It was during his physics period when he published this proof. For details, an interested reader can read my 1974 review article.[5]

The first people to use the gauge loophole were Peter Higgs,[2] Francois Englert, and Robert Brout.[6] Many of the points were later clarified by G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble.[7] I will stay within the confines of the electrodynamics of charged scalar particles for the moment so as to use the work we have already done.

We can write the Lagrangian as

$$L = -1/4(\partial/\partial x_\mu A^\upsilon(x) - (\partial/\partial x_\upsilon A^\mu(x))^2 - ((\partial/\partial x_\mu + ieA^\mu(x))\varphi^\dagger(x))(\partial/\partial x_\mu - ieA^\mu(x)\varphi(x)) + m^2\varphi^\dagger(x)\varphi(x) - 1/2(\varphi^\dagger(x)\varphi(x))^2. \quad (38)$$

The first part of the Lagrangian is the free electromagnetic part, and the last part is the bosonic part we have already seen. The middle part is the coupling term.

As before, it is convenient to split $\varphi$ into its real and imaginary parts and use the two-dimensional notation

$$\Phi = (\Phi_1, \Phi_2). \quad (39)$$

To simplify the notation, we shall introduce the $2\times 2$ matrix $q$,

$$q = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \quad (40)$$

By using the Noether techniques I described earlier, we find the conserved current,

$$J_\mu = i[(\partial/\partial x_\mu\ \Phi(x)q\Phi(x) + e\Phi(x)\cdot\Phi(x)A^\mu(x)], \quad (41)$$

whose charges generate global gauge transformations. We can now break this invariance spontaneously and make the same choice of vacuums as before. That is, we minimize the scalar boson potential as we have done in our previous examples. What is new here is the vector potential, which did not enter our previous discussion. The result is

$$\partial/\partial x^\upsilon(\partial/\partial x_\mu A^\upsilon(x) - (\partial/\partial x_\upsilon A^\mu(x)) = m^2[(1/m)\ \partial/\partial x_\mu \xi(x) - A^\mu(x)]. \quad (42)$$

By this point, the reader may wonder where is all this formalism leading? Behold, you are about to witness a miracle.

It is at this point we must choose a gauge for $A^\mu(x)$. It is convenient to use the Lorenz gauge $\partial/\partial x^\mu A^\mu(x) = 0$. Equation (42) can then be written as

$$\partial/\partial x^\upsilon[\partial/\partial x^\upsilon A^\mu(x)] = m^2[A^\mu(x) - (1/m)\ \partial/\partial x^\mu \xi(x)]. \quad (43)$$

We define a new field $B^\mu(x)$ by

$$B^\mu = A^\mu(x) - (1/m)\ \partial/\partial x^\mu \xi(x). \quad (44)$$

From the field equations, this $\xi(x)$, unlike the previous example beginning with Eq. (33), obeys the equation of an uncoupled zero mass boson—the Goldstone boson,

$$\partial/\partial x^\upsilon\ \partial/\partial x^\upsilon \xi(x) = 0. \quad (45)$$

If we make the substitution of Eq. (44) into Eq. (43) and use Eq. (45), we obtain

$$\partial/\partial x^\upsilon\ \partial/\partial x^\upsilon B^\mu = m^2 B^\mu. \quad (46)$$

The "photon" has morphed into a vector meson with mass.

Let us summarize what we have done. The electrodynamics of a charged boson with a spontaneously broken gauge symmetry in the manifestly covariant Lorenz gauge yields results consistent with a Goldstone theorem. We obtain an uncoupled massless Goldstone scalar boson $\xi$, a massive scalar boson $\eta$, and a massive vector meson $B_\mu$. Because these masses have the same origin, there is a relation between them. Because the Goldstone particle is uncoupled, it is also unobservable and can be ignored.

What happens in the Coulomb gauge where $\nabla\cdot A = 0$? I won't go though the steps but summarize the results. There is no Goldstone theorem because the gauge is not explicitly covariant and no Goldstone boson. There is a massive scalar boson $\eta$ and a massive vector meson $B_\mu$.

The first person to make full use of these ideas was Steven Weinberg in 1967.[8] To appreciate what he did, we must set the context. In 1934, Fermi produced the first modern theory of $\beta$-decay. He was an expert in quantum electrodynamics, and hence it was natural for him to use it as a template. In quantum electrodynamics the current of charged particles $J^\mu$ interacts with the electric field $A^\mu$ with a coupling of the form $J^\mu A^\mu$. Thus charged currents do not act directly with each other but only by the exchange of photons. Because there was apparently no equivalent of the photon for the weak interactions such as $\beta$-decay, Fermi directly coupled a current $J^\mu{}_N$ for the "nucleons," the neutron and proton, with a current $J^\mu{}_L$ for the "leptons," the electron and neutrino, that is, $J^\mu{}_N J^\mu{}_L$. This phenomenological theory worked very well. One could use it to calculate, for example, the energy spectrum of the electrons emitted in $\beta$-decay. But it came to seem anomalous. The "strong" interaction between nucleons, as Yukawa proposed in the prequark days, took place with the exchange of mesons, the electromagnetic interactions with photons, and presumably gravitation with gravitons. There was a suggestion of using the same meson that produced the strong interactions to produce the weak ones. This idea was abandoned. But in the 1950s, it was suggested that one or more weak heavy photons might do the trick. There were two problems. None had been observed, and the theory that was being proposed did not make any sense.

The former difficulty was easily disposed of. Because the contact theory with the currents coupled directly to each other worked well phenomenologically, it had to be that these weak mesons were very massive—too massive, it was argued, for the generation of accelerators that then existed to produce them. When they finally were produced, it turned out that their masses were about a hundred times greater than the nucleon masses. The second difficulty was qualitatively different. In the theories that were then being proposed, the weak mesons were being put in "by hand." They were just massive particles whose masses had no particular origin. If one tried to calculate anything beyond the lowest order phenomenology, we obtained terrible infinities. These infinities were much worse than those in quantum electrodynamics, which could be swept under the rug by renormalization. In short, the theory did not make any sense. Theorists were left grasping for straws. Then came Weinberg.

The "electroweak" theory of Weinberg plays on the themes we have discussed but at a higher register. The underlying Lagrangian consists of massless vector mesons, three of which—the two charged ones and the neutral one—are coupled to the scalar bosons. In addition, there is the photon, which is not coupled to these bosons. Then there are the bosons themselves, which are self-coupled as well. This Lagrangian has a global gauge symmetry, but the symmetry group is non-Abelian. In the examples that I have discussed, the effect of the global gauge transformation is to multiply the fields by a numerical phase. These examples are Abelian, and therefore it does not matter in which order two of these transformations are performed. In the non-Abelian case, it does matter. The latter case complicates the formalism but does not change the underlying methodology. Once again the gauge symmetry is broken spontaneously. The coupled vector mesons acquire masses, while the photon remains massless and there are massive scalar bosons. Apart from the fact that this method unifies two otherwise disparate interactions, it also cures the nonrenormalization issue.

There was always a sort of canary in the mineshaft interaction. It was the $\upsilon$ reaction $\nu+\vec{\nu}\rightarrow W^{+}+W^{-}$. A neutrino and an antineutrino interact and produce a pair of weak vector mesons. No one proposed to measure this reaction, but its calculation should nonetheless make sense. When this calculation is done in the conventional theory with no scalar bosons and the masses being put in by hand, the cross section increases without limit as the neutrino momentum approaches infinity. Here there are no issues of infinities caused by going to higher order, but rather there is a violation in the limit that quantum mechanics imposes on the magnitude of such cross sections due to the conservation of probability.

Weinberg observed that there is a contribution in the electroweak theory from the scalar bosons to this process, which cancels the terms that violated the quantum limit and renders the cross sections reasonable. He conjectured that the theory was renormalizable, which was proven in detail by Martinus Veltman and his student Gerhard t'Hooft.

The alert reader will notice that something is missing in this discussion. All the leptons have mass including the neutrinos, to say nothing of the masses of the neutrons and protons. What is the origin of their masses? Hopefully, the reader will indulge me in a bit of personal reminiscence. For two years in the late 1950s I was a postdoc at Harvard. Julian Schwinger was the leading light in theoretical physics at the time. We, the postdocs and junior faculty, audited whatever course he happened to be teaching. The material was always original. The lectures were on Wednesdays, and afterward the small group of us would have lunch with Schwinger at *Chez Dreyfus* in Cambridge. We would be joined by another small group from MIT that included Vikki Weisskopf. If Schwinger had any new ideas, he would try them out on Wiesskopf. As it happened on this occasion, he had developed a "theory of everything." Some of this theory survives in the work of other people. In 1962, he published a paper on "Gauge invariance and mass."[9] In it he raised the question of whether one could have a massive vector meson in a theory that had an underlying gauge invariance. This possibility is not exactly what we have been discussing, but it inspired P. W. Anderson to use these ideas in condensed matter physics.[10] Anderson used language in a nonrelativistic context, which is very similar to what we have been discussing.

I remember a lunch in which Schwinger began by saying to Weisskopf, "Now I will make you a world." The "world" was written down on a few paper napkins, one of which I saved. In any event, one of the things that he said, which has stuck with me ever since, was that scalar particles were the only ones that could have nonvanishing vacuum expectation values. He then went on to say that if you couple one of these to a fermion $\Psi$ by a coupling of the form $\Phi\bar{\Psi}\Psi$, then this vacuum expectation value would act like a fermion mass. This sort of coupling is how mass generation is done in principle for the fermions. All particles in this picture would acquire their masses from the vacuum. We are a long way from Newton.

I have avoided so far the use of the term Higgs boson—the analog in the electroweak theory of the $\eta$. Certainly, Higgs deserves the credit for first exhibiting the mechanism in the context of scalar electrodynamics. But as I have tried to show, it took other people to make it work. The Higgs boson is what is being looked for at CERN. If they find it, we shall all be happy and relieved. And if not? I am reminded of a story about Einstein. He had just received a telegram with the news that the eclipse expeditions had confirmed his general relativity prediction about the Sun bending starlight. He was very pleased with himself and showed the telegram to one of his students, Ilse Rosenthal-Schneider. She asked him what he would have done if the telegram had contained the news that the experiments disagreed with the theory. He replied, "Da könt mir halt der lieber Gott leid tun-die theorie stimmt doch. (Then I would have been sorry for the dear Lord. The theory is right)."

## ACKNOWLEDGMENTS

The author is grateful to Elihu Abrahams and Roman Jackiw for helpful communications and to an anonymous referee for a very careful reading of the paper.

[a]Electronic mail: jbernste@earthlink.net

[1]This translation from the Latin can be found in Ernst Mach, *The Science of Mechanics* (Open Court, LaSalle, IL, 1960), p. 298.

[2]P. W. Higgs, "Broken symmetries and the masses of gauge bosons," Phys. Rev. Lett. **13**, 508–509 (1964).

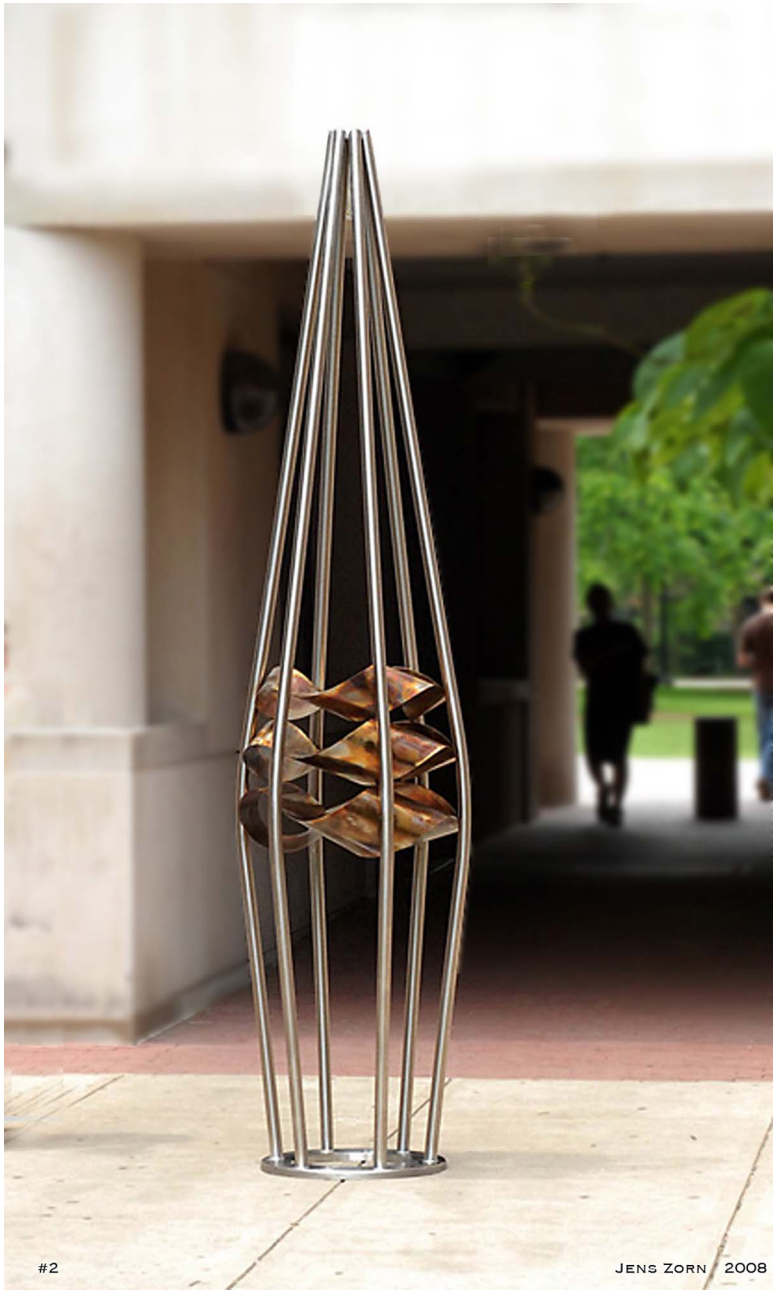
[3]This is not a misprint. Ludvig Valentin Lorenz was a 19th century Danish physicist to whom we owe this choice of gauge. I am grateful to Wolfgang Rindler for making the fact that this choice is the Lorenz gauge and not the Lorentz gauge clear so the use of "Lorenz" and not "Lorentz" is not a misprint. See J. D. Jackson, "Examples of the zeroth theorem of the history of science," Am. J. Phys. **76**, 704–729 (2008) for a brief biography of Lorenz.

[4]W. Gilbert, "Broken symmetries and massless particles," Phys. Rev. Lett. **12**, 713–714 (1964).

[5]For a discussion of the boundary question see Jeremy Bernstein, "Spontaneous, symmetry breaking, gauge theories, the Higgs mechanism and all that," Rev. Mod. Phys. **46**, 1–48 (1974).

[6]F. Englert and R. Brout, "Broken symmetries and the mass of gauge vector mesons," Phys. Rev. Lett. **13**, 321–323 (1964).

[7]G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles," Phys. Rev. Lett. **13**, 585–587 (1964).
[8]S. Weinberg, "A model of leptons," Phys. Rev. Lett. **19**, 1264–1266 (1967).
[9]J. Schwinger, "Gauge invariance and mass," Phys. Rev. **125**, 397–398 (1962).
[10]P. W. Anderson, "Plasmons, gauge invariance and mass," Phys. Rev. **130**, 439–442 (1963).




**Gravitational Lensing**

This sculpture depicts gravitational lensing by dark matter. The cause of lensing is represented here by a cluster of Möbius strips, the subtle aspects of which allude to not-yet-understood properties of dark matter. Installed at the East Lansing (MI) High School, the sculpture honors the sustained achievements in physics education by John Plough and colleagues. It is intended to remind students of the many unsolved problems in astrophysics and cosmology that remain to be attacked by new generations of scientists. This sculpture was created by Jens Zorn, Professor of Physics at the University of Michigan.