WEYL'S GAUGE INVARIANCE PRINCIPLE IN QUANTUM MECHANICS

© William O. Straub, PhD
Pasadena, California
July 23, 2005

Of all the symmetries that are expressible in quantum mechanics, perhaps the most beautiful is gauge invariance. It is completely unlike any of the other garden-variety symmetries (like translation, rotation and parity), although the symmetry it describes is mathematical rather than physical. For this reason it seems to have escaped discovery until Weyl came across it in 1918. However, his application of the gauge principle was initially based on metric geometry, a beautiful but failed idea that eventually evolved into the phase invariance principle in quantum mechanics. In the following I will sketch the basic idea behind Weyl's gauge concept, then show how it is indispensable in the derivation of the Lagrangians describing quantum electrodynamics and the Higgs mechanism.

1. Gauge Theory Basics

Gauge symmetry would absolutely have been discovered without Weyl; it was just a matter of when. Although I am not a physics historian, it's clear that the discovery of gauge symmetry paralleled the development of quantum mechanics, and the gauge concept became immediately obvious once physicists recognized the significance of the combination $\Psi^*\Psi$ as a measure of probability density. While investigating possible Lagrangians involving this quantity, it did not require any particular brilliance to see that a change in Ψ given by the *global gauge transformation*

$$\Psi' = e^{i\lambda}\,\Psi$$

where λ is an arbitrary constant, would have absolutely no effect on the Lagrangian. Indeed, Ψ' is nothing but the same wave function Ψ with an arbitrary phase factor, and both describe exactly the same physics. The term *gauge transformation* is actually a misnomer, because we're really talking about wave function *phase changes*. The "gauge" moniker is a historical carryover from Weyl's 1918 theory, which involved metric geometry, not quantum mechanics.

In quantum mechanics, the magnitude of the wave function doesn't matter; it's the "direction" that counts. To be precise, let's expand Ψ about its eigenfunctions

$$\Psi(x) = \sum a_j \psi_j(x)$$

Here, the eigenfunctions ψ play a role similar to that of ordinary base vectors in linear algebra; ψ_j represents Ψ in the "jth" direction, while a_j is a complex coefficient that represents the probability amplitude for the wave function to be found in the jth eigenstate. If Ω is any number (but not an operator), then the two quantities Ψ and $\Omega\Psi$ are considered identical – they're the same thing.

It also doesn't take any great leap of the imagination to wonder if the related transformation

$$\Psi' = e^{i\lambda(x)}\,\Psi$$

can be made without changing anything essential. This is called a *local gauge transformation*, and it is much more interesting. You can see that any operations involving differentiation of Ψ' are now going to bring down terms involving $\partial_\mu \lambda(x)$, and the Lagrangian will have to absorb these terms somehow if it is to express the same physics as Ψ. In fact, symmetries like gauge invariance have been used for some time now to deduce the proper forms of Lagrangians, whereas in the dim past physicists had to grope around looking for ones *ad hoc* to fit the experiments.

In 1918, neither Weyl nor anyone else had the faintest clue about the phase invariance of the wave function because the wave function hadn't been discovered yet. Instead, Weyl was investigating gauge transformations of the metric tensor

$$g'_{\mu\nu} = e^{\lambda(x)} g_{\mu\nu}$$

and the effect of these transformations on Riemannian and non-Riemannian geometry. Weyl noticed that the magnitude of an arbitrary vector ξ^μ would undergo a rescaling under a gauge transformation given by

$$\begin{aligned} |\xi|^2 &= g_{\mu\nu}\xi^\mu\xi^\nu, \\ |\xi'|^2 &= g'_{\mu\nu}\xi^\mu\xi^\nu \\ &= e^{\lambda(x)}|\xi|^2 \end{aligned}$$

and he wondered if such a regauging would alter any essential physics. In essence, what Weyl was thinking was that nature might not care if vector magnitude were rescaled from one spacetime point to another and, taking this one step further, he thought that nature might actually *require* gauge invariance of metric spacetime. As you will see, this idea closely resembles the arbitrariness of wave function magnitude and the very profound fact that its arbitrariness is *demanded* if electrodynamics is to be part of the formalism. Weyl went on to discover (or invent) a spacetime manifold that was consistent with this idea, motivated by the much greater prospect that it would lead to a unification of electrodynamics and gravitation. Weyl failed in this endeavor, but in 1929 he hit the jackpot when he applied the gauge principle to the newfangled wave function of quantum theory.

When Weyl first proposed quantum mechanical gauge invariance in 1929, he did not know that there are forces other than electromagnetism and gravity at work in nature. However, it is comforting to know that before his death in 1955, he was well aware of the strong and weak forces and that they might be describable by a gauge theory. Today, the gauge principle is arguably the most powerful concept in all of modern physics. It underlies all of the Yang-Mills theories and is a key component in string theory and its more recent variant, M theory.

Nevertheless, one cannot but feel a trace of regret on the part of Weyl in his 1929 paper. By the paper's end, he admits that the gauge principle applies not to gravity but to quantum mechanics, but there is a tinge of sadness that things had to work out this way (or maybe it's just my imagination). After all, the free-space equations of electrodynamics can be derived from the gauge-invariant Lagrangian

$$I = \int \sqrt{-g} F_{\mu\nu} F^{\mu\nu} d^4x$$

which is quadratic in the electromagnetic field tensor $F_{\mu\nu}$. By comparison, the Lagrangian for Einstein's free-space gravity equations is

$$I = \int \sqrt{-g} R \, d^4x$$

which is neither gauge-invariant (in the sense that Weyl envisioned) nor quadratic in the integrand. In 1918, Weyl proposed a Lagrangian that included the square of the Ricci scalar, R^2, but neither this nor the scalar $R_{\mu\nu}R^{\mu\nu}$ went anywhere, and the theory languished for several years until Einstein himself pounded the last nail into its coffin around 1921. It was a beautiful idea that seemed to have no application in modern physics.

2. Background

Weyl's gauge idea is closely related to something called Noether's Theorem. You've probably heard of it, but it's such a beautiful concept that I'm going to briefly describe both the theorem and the remarkable woman who came up with it – Emmy Noether.

Since gauge invariance and Noether's theorem presuppose some knowledge of extremal principles in variational calculus, you should also have a basic understanding of the mathematical concepts behind *Lagrangians* and *Hamilton's Principle*. I'm not going to discuss these in any detail, but I'll lay out the basics along the way.

3. Emmy Noether

Amalie (Emmy) Noether (pronounced *nuhr′ter*, somewhat like *Goethe*) was born in Erlangen, Germany on March 23, 1882, the daughter of Max and Ida Noether. Father Max was a well-known mathematician at the University of Erlangen, and he undoubtedly influenced Emmy's decision to forget about her earlier interests (linguistics and dancing) and take up mathematics. This was no small feat, as German women at the time were discriminated against in math and science. But she persevered, and by a process of formal and informal class attendance at the university she managed to earn a PhD in mathematics in 1907 (she was the only doctoral student of the noted mathematician Paul Gordan who, with Rudolph Clebsch, discovered the Clebsch-Gordan coefficients). Since she was not allowed to teach, she helped her father with his studies and began publishing papers on her own. In spite of the constant professional discrimination she experienced, she quickly developed a reputation in Germany as a brilliant algebraist. The noted mathematicians Weyl, Hilbert and Klein soon took notice of her, and in 1915 Hilbert was able to secure an unpaid position for Noether at the University of Göttingen. In her first few years at the school she worked with Hilbert and others on Einstein's gravitation theory. During this time she hit upon a connection between integral invariants and conservation laws, which she published in July 1918. Her discovery, which is rightly called *Noether's Theorem*, is remarkable in terms of its power and simplicity.

Emmy Noether, 1882-1935

Noether's dedicated teaching style and caring attitude made her enormously popular at Göttingen, and her predominantly male students adored her (Hilbert referred to them as *Noether Jungen*, or Noether's boys). However, in spite of her genius Noether had several things going against her. For one, she could not avoid discrimination by the legitimate mathematical community because of her sex. She was also a Jew, and a pacifist Jew at that, and when the Nazis took power in 1933, Noether was summarily fired (along with every other Jewish professor in the country). She taught in secret from her apartment for a few months, and then was offered a teaching position at Bryn Mawr College in Pennsylvania. By accepting, she joined a long list of Jewish intellectuals expatriating from Germany to more hospitable climes.

Noether was deeply dedicated to mathematics and teaching, and as a result she ignored her health. She became overweight, and pictures of her in later years show that she had completely given up caring about her physical appearance. She taught at Bryn Mawr until her untimely death at age 53 on April 14, 1935, believed to be the result of a post-operative infection following removal of a large ovarian tumor. Upon hearing of her passing, Einstein wrote

> *In the judgment of the most competent living mathematicians, Fräulein Noether was the most significant creative mathematical genius thus far produced since the higher education of women began.*

Weyl himself travelled to Bryn Mawr to give the memorial address, where he noted that

> *She was not clay, pressed by the artistic hands of God into a harmonious form, but rather a chunk of human primary rock into which He had blown His creative breath of life.*

Weyl and Noether had been close friends and colleagues for twenty years (he often good-naturedly referred to her using the masculine title *der Noether* in honor of her mathematical abilities), and he was heartbroken at her passing. In his address he decried the prejudice she had experienced as a female mathematician and her travails as a Jewish intellectual. He recounted fond memories from their teaching days together at Göttingen, and confessed sincerely that he considered her to be his superior as a mathematician. Quite an admission considering Weyl's own abilities!

Now that you know a little about the woman, let's look at her theorem.

4. Noether's Theorem

It has been known since Hamilton's day that the dynamics of a classical physical system can be obtained by extremalizing (usually minimizing) the integral quantity

$$I = \int L\, dt$$

where L is the Lagrangian density of the system. In classical physics, the Lagrangian is just the difference between the kinetic and potential energies. In general relativity, the integral takes on the form of the four-dimensional invariant scalar quantity

$$I = \int \sqrt{-g}\, L\, dx,$$

(in quantum mechanics and field theory the metric determinant term is usually suppressed.) The Lagrangian is most often some function of the four spacetime coordinates and their first derivatives, $L = L(x, \partial_\mu x)$,or some field, $L = L\left[\phi(x), \partial_\mu \phi(x)\right]$. In quantum field theory, the Lagrangian retains its composition of kinetic and interaction terms, but these terms do not have the outward appearance of kinetic and potential energies (the interaction term often includes a mass term as well). For example, the Lagrangian of a self-interacting scalar field $\phi(x)$ (for a particle like a boson) might appear as

$$L = \frac{1}{2}\partial_\mu \phi\, \partial^\mu \phi - \frac{1}{2} m^2 \phi^2 - \frac{1}{4}\lambda \phi^4 \tag{4.1}$$

where λ is a coupling constant that determines the strength of the interaction of the particle with itself.

Deriving Noether's Theorem is simplicity itself. Note that any variation of the Lagrangian can be written as

$$\delta L = \frac{\partial L}{\partial x}\, \delta x + \frac{\partial L}{\partial(\partial_\mu x)}\, \delta(\partial_\mu x) \tag{4.2}$$

If we integrate this quantity over all space, we get

$$\begin{aligned} \delta I &= \int \delta L\, dx \\ &= \int \left[\frac{\partial L}{\partial x}\, \delta x + \frac{\partial L}{\partial(\partial_\mu x)}\, \delta(\partial_\mu x) \right] dx \\ &= \int \left\{ \frac{\partial L}{\partial x}\, \delta x - \partial_\mu \left[\frac{\partial L}{\partial(\partial_\mu x)} \right] \right\} \delta(x)\, dx \end{aligned} \tag{4.3}$$

where I have integrated by parts in the last integral. Now, if the system in question does not change under the variation, then we say that the system has a *symmetry* associated with that particular variation. This results by setting (4.2) to zero, from which we get

$$\frac{\partial L}{\partial x}\, \delta x = -\frac{\partial L}{\partial(\partial_\mu x)}\, \delta(\partial_\mu x)$$

Plugging this into the integrand in (4.3) and setting it to zero as well, we then have

$$\begin{aligned} \frac{\partial L}{\partial(\partial_\mu x)}\, \delta(\partial_\mu x) + \partial_\mu \left[\frac{\partial L}{\partial(\partial_\mu x)} \right] \delta(x) &= 0, \quad \text{or} \\ \partial_\mu \left[\frac{\partial L}{\partial(\partial_\mu x)}\, \delta(x) \right] &= 0 \end{aligned} \tag{4.4}$$

For a field $\phi(x)$, this goes like

$$\partial_\mu \left[\frac{\partial L}{\partial(\partial_\mu \phi)} \, \delta\phi \right] = 0 \tag{4.5}$$

If we identify the bracketed quantity in either (4.4) and (4.5) as a four-density J^μ, then we have the usual conservation condition $\partial_\mu J^\mu = 0$. Note that the variations $\delta(x)$ and $\delta\phi$ may be completely arbitrary, so we can say that if a dynamical system has a symmetry of some kind, then there is a corresponding conservation law associated with that symmetry. This is Noether's Theorem – it took more space to describe the woman than it did her theorem!

This is the power of Noether's Theorem – every conservation law corresponds to a particular symmetry in the Lagrangian (and vice versa, with qualifications). This is pretty much the way physicists have approached Lagrangians since the time she wrote it down – start with a scalar integral expression, get the kinetic terms in, then the interaction terms, postulate some kind of symmetry, then tinker with the expression until it is invariant with respect to that symmetry. The development of the Standard Model of particle physics – the most accurate physical theory mankind has developed to date – reflects this approach.

I do not know what Noether (a secular Jew) personally thought of her theorem, but I would like to think that she saw God's hand in it. It is simply too beautiful and profound for God to have overlooked it as a working principle when he designed the universe. When it popped into Noether's mind, I can almost imagine God thinking to himself "Aha! Now they're really on to something interesting down there!"

5. The Lagrangian of Quantum Electrodynamics

Armed with these simple but extremely powerful mathematical tools, it is now a simple matter to derive the Lagrangian of quantum electrodynamics using the gauge invariance concept. To do this, we start with the free-particle Lagrangian for Dirac's equation, which is

$$L = i\hbar c \overline{\Psi} \gamma^\mu \, \partial_\mu \Psi - mc^2 \overline{\Psi} \Psi \tag{5.1}$$

where $\Psi(x)$ is the Dirac field for a spin one-half particle and $\overline{\Psi}(x) = \Psi(x)\gamma^0$ is the adjoint spinor. Notice that the leading term represents the kinetic contribution, while the second term brings in the particle's mass; there's no interaction term, because we're only considering a free particle. The corresponding action quantity is then

$$S = \int L \, dx = \int \left[i\hbar c \overline{\Psi} \gamma^\mu \, \partial_\mu \Psi - mc^2 \overline{\Psi} \Psi \right] \, dx$$

It is obvious that performing the variation $\delta\Psi^\dagger$ on the Dirac action gives the Dirac equation, $[i\hbar\gamma^\mu \partial_\mu - mc] \, \Psi = 0$. Conversely, taking the variation $\delta\Psi$ is almost as easy (you have to do an integration by parts), and you then get the adjoint Dirac equation, $i\hbar\partial_\mu \overline{\Psi} \gamma^\mu + mc\overline{\Psi} = 0$. So far so good.

Now let's see what happens when we demand that the Dirac equation be invariant with respect to a gauge transformation of the kind

$$\Psi' = e^{i\lambda} \Psi$$

where λ is some arbitrary, non-zero constant (this is called a *global* gauge transformation because the it applies the same factor to the field everywhere in space). In what follows, it will suffice to concern ourselves only with infinitesimal transformations, so we can write, to first order, $\Psi' = e^{i\lambda}\Psi = (1 + i\lambda)\Psi$. The gauge variation can then be expressed as $\delta\Psi = \Psi' - \Psi = i\lambda\Psi$. Similarly, we have $\delta\overline{\Psi} = -i\lambda\overline{\Psi}$. Obviously, the quantity $\overline{\Psi}\Psi$ is invariant; also, since the Dirac matrices and λ are constants, it is easy to see that $\delta L = 0$. Consequently, the free-particle Dirac Lagrangian is invariant with respect to global gauge transformations.

You might be tempted to think that this is no big deal, but if you go back to Noether's equation (4.5) you can see that even with a constant gauge factor there is a conserved quantity. Plugging $\delta\Psi = i\lambda\Psi$ into (5.1), we see that this quantity is the Lorentz vector $\overline{\Psi}\gamma^\mu\Psi$. It is conventional to tack the electronic charge e onto the conservation expression, so that $\partial_\mu(e\overline{\Psi}\gamma^\mu\Psi) = 0$. The conservation of electric charge is thus a consequence of the global gauge symmetry of the Dirac Lagrangian.

It is inevitable that the constant gauge parameter λ should be elevated to a function of space, $\lambda(x)$. How does this change things? From the Dirac Lagrangian, it is easy to see that δL will now involve the term $\partial_\mu\lambda$. In order to preserve the condition $\delta L = 0$, we must tinker with the Lagrangian to keep it invariant with respect to what is now called a *local gauge transformation*, $\delta\Psi = i\lambda(x)\Psi$.

The simplest approach is just to add a term to the Lagrangian that has a lower index like ∂_μ, so we write

$$L = i\hbar c\overline{\Psi}\gamma^\mu\,\partial_\mu\Psi + e\overline{\Psi}\gamma^\mu A_\mu\Psi - mc^2\overline{\Psi}\Psi \tag{5.2}$$

where $A_\mu(x)$ is a new field quantity. Its association with the Dirac spinors means that it should be viewed as an interaction between the fields, where the charge e acts like a coupling constant. If we now pass the gauge variation operator through (5.2), we get

$$\begin{aligned}\delta L &= i\hbar c(-i\lambda\overline{\Psi})\gamma^\mu\,\partial_\mu\Psi + i\hbar c\overline{\Psi}\gamma^\mu\,\partial_\mu(i\lambda\Psi) + e(-i\lambda\overline{\Psi})\gamma^\mu A_\mu\Psi \\ &\quad + e\overline{\Psi}\gamma^\mu\,(\delta A_\mu)\,\Psi + e\overline{\Psi}\gamma^\mu A_\mu(i\lambda\Psi)\end{aligned}$$

Collapsing terms, we're left with

$$\delta L = \overline{\Psi}\gamma^\mu\Psi\,(e\delta A_\mu - \hbar c\partial_\mu\lambda)$$

This clearly vanishes if the field A_μ varies under a local gauge variation as

$$\delta A_\mu = \frac{\hbar c}{e}\,\partial_\mu\lambda \tag{5.3}$$

Of course, this is precisely how the electromagnetic four-vector varies under a gauge variation. In view of this, we tentatively identify A_μ as the four-potential of electrodynamics. The Lagrangian (5.2) shows how this field couples with the Dirac field Ψ.

However, we're not quite finished. Recall that Lagrangians tend to be composed of kinetic and potential (or interaction) terms, like $L = T - V$ in classical mechanics. Also, in quantum field theory, there often is a mass term thrown in as well. The kinetic term in the Dirac Lagrangian is $\overline{\Psi}\gamma^\mu\,\partial_\mu\Psi$, and the interaction term is $e\overline{\Psi}\gamma^\mu A_\mu\Psi$ (which also defines the interaction term for the field A_μ). The field Ψ also has a mass term, $mc^2\overline{\Psi}\Psi$. But where are these terms for the A_μ field?

Let's tackle the mass term first. Such a term for the field A_μ would have to be a scalar, and the only one that might work is something proportional to the quadratic quantity $mA_\mu A^\mu$. But as you can see for yourself, the gauge variation of $\delta(A_\mu A^\mu)$ is *non-zero*. We are forced to conclude that the field A_μ is massless. But that's just fine, because it again justifies its identification with the electromagnetic field (photons), which of course is massless.

To get a kinetic term for A_μ into the Lagrangian, we make a straightforward appeal to the standard Lagrangian of electrodynamics, which is

$$L = J^\mu A_\mu + \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

where $F_{\mu\nu}$ is the antisymmetric stress-energy tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

and J^μ is the source term. The term $F_{\mu\nu}F^{\mu\nu}$also has the nice property that it is invariant with respect to gauge transformations. Notice also that by identifying the Lorentz vector $\overline{\Psi}\gamma^\mu\Psi$ with the source vector J^μ, we arrive at yet another suggestive connection between the Dirac Lagrangian and that of electrodynamics. You now know the famous Lagrangian for quantum electrodynamics:

$$\begin{aligned}L &= i\hbar c\overline{\Psi}\gamma^\mu\,\partial_\mu\Psi + e\overline{\Psi}\gamma^\mu A_\mu\Psi - mc^2\overline{\Psi}\Psi + \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &= i\hbar c\overline{\Psi}\gamma^\mu\left[\partial_\mu - \frac{ie}{\hbar c}A_\mu\right]\Psi - mc^2\overline{\Psi}\Psi + \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \qquad (5.4)\end{aligned}$$

which describes the behavior of an electron (actually, any spin-1/2 fermion carrying charge e) in the presence of an electromagnetic field. The term in brackets is often abbreviated as

$$\mathcal{D}_\mu = \partial_\mu - \frac{ie}{\hbar c}\,A_\mu \tag{5.5}$$

and is known as the *covariant derivative operator* (not to be mistaken with the covariant derivative of differential geometry). The covariant derivative is a standard trick in quantum field theory, and it works not only for one-dimensional unitary gauge transformations (like the one we have been using), but for higher-dimensional ones as well. When we want a gauge-invariant expression, we just replace the partial derivatives with their covariant counterparts and tack on the $F_{\mu\nu}F^{\mu\nu}$ term.

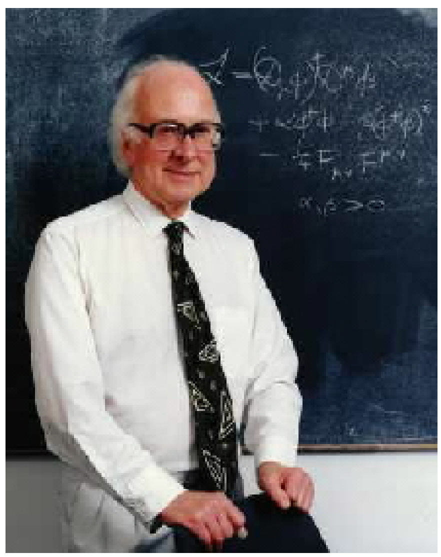
Note, however, that the above Lagrangian does not describe the interactions of bunches of electrons or their interaction with other particles, like muons or protons. That little problem requires quantum field theory.

And it wouldn't have been possible without Weyl!

6. The Higgs Mechanism

In the Lagrangian for quantum electrodynamics, we showed that gauge invariance required the mass term for the electromagnetic field A_μ to vanish. This made sense, because the carriers of the electromagnetic force, photons, are massless bosons of spin one. By comparison, the carriers of the strong nuclear force, gluons, are also massless bosons, while the weak force, which is responsible for certain kinds of radioactive decay, is also described by bosons (with zero spin). But these force carriers (the $\mathrm{W}^{\pm}$ and Z^0 particles) have significant masses (around 80-90 MeV). Now, bosonic quantum field theory requires a Lagrangian similar to that given by (4.1). If gauge invariance is to be considered a fundamental symmetry of all interactions, how can these bosons have mass, since mass destroys gauge invariance?

In 1961, the British theoretical physicist Peter Higgs of Edinburgh University considered this problem, and wondered if there was any way that massive particles could be described by a scalar (spin zero) Lagrangian that remains gauge invariant. He discovered that there is indeed a way, and his discovery has since become famous even though it has not yet been experimentally verified.



Peter Higgs

Consider again the bosonic Lagrangian

$$L = \frac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - \frac{1}{2}m^2\phi^2 - \frac{1}{4}\lambda\phi^4 \tag{6.1}$$

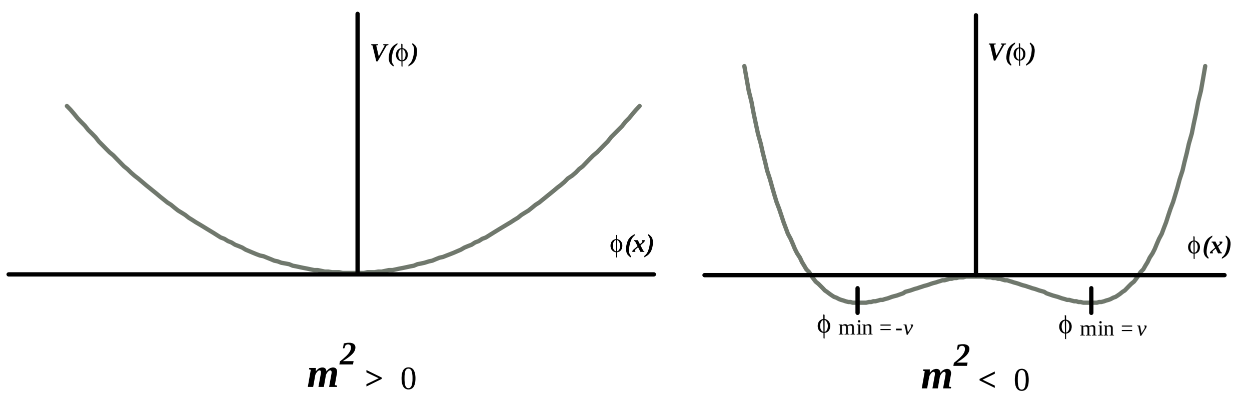
We now ask what the minimum energy point (the value of the field for minimum energy) is for the interaction term $V(\phi) = \frac{1}{2}m^2\phi^2 + \frac{1}{4}\lambda\phi^4$. An awareness of this point is critical, because invariably a perturbative approach must be used to solve the problem (for an explanation, see my other write-up on quantum field theory), and perturbations are always developed around the point of minimum energy. Taking the derivative and setting it to zero, we have

$$\begin{aligned}\frac{\partial V}{\partial\phi} &= m^2\phi + \lambda\phi^3 \\ &= \phi(m^2 + \lambda\phi^2) = 0\end{aligned}$$

Obviously, one solution is $\phi_{\min} = 0$. However, there are two more:

$$\phi_{\min} = \pm\sqrt{\frac{-m^2}{\lambda}}$$

At first glance this is meaningless, because these minimal points are pure imaginary, making V a complex quantity. But there seems to be a couple of ways around this. One is to assume that the self-coupling constant is negative; the other is to have a negative mass term m^2 (which would make the particle mass imaginary). However, for physical reasons λ must have the same sign as the kinetic term, so that option is out. The second, imaginary mass, seems to be even more bizarre. How can you have imaginary mass?



The answer involves our expression for the field ϕ. As you will shortly see, by rewriting the form of ϕ and coupling it with an imaginary mass, will give us a Lagrangian having the proper mass sign (an even better approach is to treat the m^2 in (6.1) as just a *parameter* rather than a mass term). So let's go ahead and write the Lagrangian with the "wrong" mass signature:

$$L = \frac{1}{2}\partial_\mu \phi\, \partial^\mu \phi + \frac{1}{2}m^2\phi^2 - \frac{1}{4}\lambda\phi^4 \tag{6.2}$$

which now gives

$$\phi_{\min} = \pm\frac{m}{\sqrt{\lambda}}$$

Keep in mind two things about this quantity: one, it is symmetric with regard to space reflection; that is, $L(\phi) = L(-\phi)$; and two, it is invariant with respect to global gauge transformations. Let us now define a new (real) field $\eta(x)$ using the substitution

$$\phi(x) = \eta(x) + v$$

where $v = m\lambda^{-1/2}$. Unlike ϕ, the new field $\eta(x)$ has the advantage of including the point of minimum energy. Plugging this into (6.2), we get the modified Lagrangian

$$L = \frac{1}{2}\partial_\mu \eta\, \partial^\mu \eta - m^2\eta^2 - \frac{1}{4}\lambda\eta^4 - v\lambda\eta^3 + \frac{1}{4}v^4\lambda$$

This quantity now has the correct mass signature and two self-couplings (the η^3 and η^4 terms). The last term is a constant and is irrelevant. Notice, however that this Lagrangian has lost its reflection symmetry and is no longer gauge invariant. So what has been gained?

The main point of this exercise is to demonstrate that a perturbation approach would not converge for the original Lagrangian (6.1) because the true energy minimum cannot be reached for any order of the perturbative expansion, thus dooming the approximation effort. Secondly, it shows that by allowing the true energy minimum to enter the calculation, space symmetry is lost; that is, the Lagrangian has undergone what is called *spontaneously broken symmetry*, a rather technical issue that has tremendously important implications in field theory. And thirdly, it shows that the mass term can have the “wrong” sign and everything can be made right again (if you overlook the additional terms, which represent higher-order interactions). And, in case you haven’t already noticed, the two Lagrangians describe absolutely the *same* physics. Nothing has changed; it’s just that the revised Lagrangian is suitable for perturbative expansion, while the original is not.

The concept of spontaneously broken symmetry can be visualized by considering the analogy of a vertical column in structural engineering that is pinned at both ends. The column carries a load that is slowly being increased. At first everything is symmetrical, but when the load reaches a certain critical point, the column suddenly buckles outward. In its first failure mode, the column just bends outward along some vertical plane whose direction is completely random. Symmetry is lost, but this is in response to the system adjusting to a non-symmetric point of equilibrium.

The model Lagrangian we have just examined is limited in the sense that there are only two discrete energy minima at the field points $\phi_{\min} = \pm v$ (see above figure). A more realistic model would have a continuum of minima, and this can be achieved by considering a Lagrangian involving two fields, ϕ_1 and ϕ_2, which can be combined as the complex field

$$\phi(x) = \phi_1(x) + i\phi_2(x)$$

(imagine the second graph in the figure being rotated about the origin). Our starting Lagrangian will now look like

$$L = \frac{1}{2}\partial_\mu\phi^*\,\partial^\mu\phi + \frac{1}{2}m^2\phi^*\phi - \frac{1}{4}\lambda(\phi^*\phi)^2 \tag{6.3}$$

Notice that this quantity is also invariant with respect to space reflection and global gauge transformations. However, it should not be surprising that by having a two-component field, we will have to introduce *two* new fields in the description of the complex ϕ-field. This is the price one has to pay to make perturbation a practical approach. However, the theoretical consequences of this apparent distraction are nothing short of fantastic.

The interaction term in the Lagrangian (6.3) can be expanded in terms of the individual fields to give

$$V = -\frac{1}{2}m^2(\phi_1^2 + \phi_2^2) + \frac{1}{4}\lambda(\phi_1^2 + \phi_2^2)^2$$

The energy minima are now continuous and lie in the circular plane described by

$$\phi_{1\min}^2 + \phi_{2\min}^2 = \frac{m^2}{\lambda}$$

We could now write the complex field $\phi(x)$ in terms of two new real fields $\eta(x)$ and $\xi(x)$ and the energy minimum m^2/λ as something like

$$\phi = \eta + v + i\xi$$

(where $v^2 = m^2/\lambda$), but there is another expression that will vastly simplify the calculations when we plug this expression for ϕ into the Lagrangian. I saw it first in Halzen and Martin’s text; it’s

$$\begin{aligned} \phi &= (\eta + v)e^{i\xi/v}, \\ \phi^* &= (\eta + v)e^{-i\xi/v} \end{aligned} \tag{6.4}$$

which is essentially the polar form of $\eta + v + i\xi$. Inserting this into our Lagrangian, we get, after some reduction,

$$L = \frac{1}{2}\left[\partial_\mu\eta\,\partial^\mu\eta + \partial_\mu\xi\,\partial^\mu\xi\right] - m^2\eta^2 - \lambda v\eta^3 - \frac{1}{4}\lambda\eta^4 + \frac{1}{4}\lambda v^4 + \frac{\lambda}{2m^2}\left(\partial_\mu\xi\,\partial^\mu\xi\right)\eta^2 + \frac{1}{v}\left(\partial_\mu\xi\,\partial^\mu\xi\right)\eta \tag{6.5}$$

The field $\eta(x)$ looks fine; it has a kinetic term, a mass term with the right sign, and a few couplings with the $\xi(x)$ field (admittedly, the last two look odd, but they’re just couplings). We interpret the field $\eta(x)$ as a massive spin-zero boson. By the same logic, the $\xi(x)$ field must describe a massless boson of zero spin (it's called a *Goldstone boson* in the trade). Unfortunately, *there is no such particle in nature!*

This is indeed a disaster, and it would seem that all of this has been a colossal waste of time. But not to fear, as we have one more trick up our sleeve – Weyl's *local gauge symmetry*. Let's replace the partial derivatives in (6.3) with the covariant derivatives of the previous section (Equation 5.5) and add the $F_{\mu\nu}F^{\mu\nu}$ term. The Lagrangian will then look like

$$L = \frac{1}{2}\left[\left(\partial_\mu - \frac{ie}{\hbar c}A_\mu\right)(\eta+v)e^{-i\xi/v}\left(\partial^\mu + \frac{ie}{\hbar c}A^\mu\right)(\eta+v)e^{i\xi/v}\right] + \frac{1}{2}m^2(\eta+v)^2 + \frac{1}{4}\lambda(\eta+v)^4 + \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

Grinding this out is straightforward but a bit tedious (although the exponential forms in (6.4) really help out here), and we end up with

$$\begin{aligned} L &= \frac{1}{2}\partial_\mu\eta\,\partial^\mu\eta - m^2\eta^2 - v\lambda\eta^3 - \frac{1}{4}\lambda\eta^4 + \frac{1}{4}\lambda v^4 + \frac{e^2}{2\hbar^2c^2}(\eta+v)^2 A_\mu A^\mu + \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &\quad + \frac{1}{2v^2}(\eta+v)^2\,\partial_\mu\xi\,\partial^\mu\xi + \frac{e}{\hbar c v}A_\mu\partial^\mu\xi \end{aligned} \tag{6.6}$$

This expression still contains the massless Goldstone boson $\xi(x)$. But wait! Remembering (5.3), we are still free to specify an arbitrary gauge for the A_μ field. If we pick the gauge

$$A'_\mu = A_\mu - \frac{\hbar c}{2ev}\,\partial_\mu\xi$$

then the last two terms in (6.6) vanish completely, and the Lagrangian goes over to

$$\begin{aligned} L &= \frac{1}{2}\partial_\mu\eta\,\partial^\mu\eta - m^2\eta^2 - \lambda v\eta^3 - \frac{1}{4}\lambda\eta^4 + \frac{e^2m^2}{2\hbar^2c^2\lambda}A_\mu A^\mu + \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &\quad + \frac{e^2}{2\hbar^2c^2}\left(A_\mu A^\mu\right)\eta^2 + \frac{e^2 v}{\hbar^2c^2}\left(A_\mu A^\mu\right)\eta + \frac{1}{4}\lambda v^4 \end{aligned}$$

You might not have noticed, but a miracle has just occurred – not only has the massless boson ξ disappeared, but the photonic gauge field A_μ has acquired a mass ($\sqrt{e^2m^2/2\hbar^2c^2\lambda}$). In physics parlance, we say that the gauge field A_μ has "eaten" the Goldstone boson and gained weight!

This little lesson in mathematical chicanery is known as the *Higgs mechanism*. It allows us to assign non-zero masses to the electroweak bosons $W^\pm$ and Z^0 while maintaining the gauge invariance of the Lagrangian, at the cost of introducing a new field $\eta(x)$, which is known as the Higgs field (mediated by the *Higgs boson*, which some physicists have dubbed the *God Particle*). To date, the Higgs boson has not been observed. Its mass, estimated to be in the range of 117 GeV to 251 GeV, has so far escaped detection because existing particle accelerators are not powerful enough to elucidate such a massive particle. However, the Large Hadron Collider, currently under construction at CERN in Switzerland, will almost certainly detect the Higgs, if it exists. The LHC is scheduled to begin operation in 2007.

Physicists believe the Higgs field is behind the mystery of particle mass. Why is a gluon massless, while an electron has a mass of 0.51 MeV? It is thought that there is a universal Higgs field or "ether" permeating the vacuum that is able to interact with particles and slow down their movement, thus giving them mass. A useful analogy is sometimes given of a famous movie star who, though wearing dark glasses and wishing to be left alone, enters a room and is instantly recognized and surrounded by autograph-hungry fans, who slow her down. The Higgs field supposedly interacts with particles in much this way and, by slowing them down, effectively gives them inertial mass.

Symmetry – Evidence of a Creator?

You are probably aware that the natural world is neither good nor evil, moral or immoral; these labels simply do not apply to nature. A pack of hyenas ripping apart a baby zebra is no more "bad" than a rainbow is "good" in any sense of the word – unfeeling nature simply does not care. By this same reasoning, an accidental universe, bereft of a Creator, would have no use for any kind of mathematical or physical symmetry principle; any chaotic arrangement would do nicely, and the creatures living in such a universe (assuming their existence were even possible) would certainly not care that symmetry principles didn't exist. However, we know without question that such symmetries exist in nature and that they offer a profound way of understanding how things work. We humans look upon this and see order and structure, which to us is a type of beauty. Since only intelligent animals can be truly aware of beauty, it is logical to assume that humans are a notch higher than the other animals (although I sometimes wonder about this). Therefore, the existence of mathematical symmetry in nature must point to something profound. Scientists like Einstein, Dirac and Weyl were intimately acquainted with this beauty, and their minds were deeply moved by it. Dirac once went so far as to remark that beauty in a mathematical equation is more important than its ability to reproduce experimental results.

In his excellent (but to my mind, wrong) best-selling 1986 book *The Blind Watchmaker*, author Richard Dawkins argues that nature need not have an intelligent designer working behind the scenes because processes like evolution, which itself is basically a long-term stochastic process, can produce any level of complexity in living systems if given some raw ingredients and a long enough period of time to work. In Dawkins' view, the ticking gold watch found at the seashore does not require a watchmaker, just enough matter and time to put itself together. Statistically and probabilistically, he is correct – quantum-mechanically, there is a small but finite probability that a complex system will simply materialize out of nothingness, just as virtual particles can pop into and out of existence in the vicinity of a charged particle. It's the same as saying that if you toss a few million bits of paper into the air, each containing a single Hindu character, eventually they will land on the ground to form the *Baghavad Gita*, in iambic pentameter, no less.

However, the probability that such macro-processes will occur, while not precisely zero, is unimaginably small. For example, the radioactive decay rates of certain nuclei (bismuth-209 and several lead isotopes) may be so small that their half-lives exceed the known age of the universe by many orders of magnitude. Dawkins touches on this but neglects to consider the quantum-mechanical implications in terms of actual human experience (for example, the *Copenhagen Interpretation* of QM basically says that if you don't actually observe something, then you don't have the right to draw any logical inferences about its existence or behavior). He also does not mention the fact that, as improbable as such an event might be, when it does happen, it could happen *anywhere* in the universe, not necessarily on earth, making its discovery exceedingly unlikely in the tiny corner of the universe we happen to live in. That is, if a gold watch were to suddenly spring into existence, it would almost certainly not do so on this insignificant planet. He also does not address the fact that in a strictly evolutionary Dawkins-world there would be no need for mathematical symmetry, which is absolutely necessary for the watch to exist, much less function. And perhaps most importantly, he cannot explain the fact that uncaring, statistical evolution must have a driving force to make things "go," but evolution by itself cannot provide the driving force.

What is the driving force? Ah, that is the great mystery of life that no one yet understands! Religious leaders can say only that life is "God's will," which is something of a cop-out, but the evolutionists have absolutely no answer, because unthinking nature does not have the "will" to produce life. It seems far simpler to me just to assume the existence of a God from the outset; whether it's Jehovah, Jesus, Allah or Brahma is strictly a matter of preference or faith. To my way of thinking, if God had done nothing more in the beginning than create a crude, solitary single-cell microorganism with the innate driving force to reproduce itself and evolve over time, it would in no way detract from my opinion of him, and he would be just as worthy of praise. I believe in evolution, but I believe it is just one of God's tools that allows his creation to adapt to changing environmental conditions. The Old Testament says that God created man from dust, but it does not specify exactly how he did this or how long it took. Adam and Eve may have been nothing more than a couple of highly-evolved *australopithecines* whom God had endowed with intelligence and a soul. If you're a strict literalist, and believe God's days were exactly 24 (integer) hours long in the beginning and that Eve sprung from Adam's rib bone, then there are any number of illogical allegories in the Old and New Testaments that you'll need to explain to me before you even begin to start in on evolution.

So what do I conclude from all this raving? That God exists because mathematical symmetry, necessary for all order and beauty in the world, exists and can be recognized and appreciated by human beings. And I further believe that Weyl's gauge symmetry, in recognition of its indispensable role in the development of modern physics, is the most sublime of all God's symmetries.

References

1. D. Griffiths, *Introduction to Elementary Particles*. Wiley & Sons, New York, 1987.
2. F. Halzen and A. Martin, *Quarks and Leptons*: *An Introductory Course in Modern Particle Physics*. Wiley & Sons, New York, 1984.
3. L. Ryder, *Quantum Field Theory*. Cambridge University Press, 2nd ed., Cambridge, 1996.