PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Review

Cite this article: Kibble TWB. 2015 Spontaneous symmetry breaking in gauge theories. *Phil. Trans. R. Soc. A* **373**: 20140033. http://dx.doi.org/10.1098/rsta.2014.0033

One contribution of 12 to a Discussion Meeting Issue 'Before, behind and beyond the discovery of the Higgs boson'.

Subject Areas:
high-energy physics, particle physics

Keywords:
symmetry breaking, gauge theory, electroweak theory

Author for correspondence:
T. W. B. Kibble
e-mail: kibble@ic.ac.uk

Spontaneous symmetry breaking in gauge theories

T. W. B. Kibble

Blackett Laboratory, Imperial College London, London SW7 2AZ, UK

The aim of this historical article is to describe the development of the idea of spontaneous symmetry breaking in gauge theories as seen from my perspective as a member of Abdus Salam's group at Imperial College London, UK. Beginning with an account of particle physics in the years after the Second World War, I describe early attempts at constructing a unified theory of weak and electromagnetic interactions, the obstacles encountered and how they were eventually overcome with the mass-generating mechanism incorporating the idea of spontaneous symmetry breaking, one of whose features is the now-famous Higgs boson.

1. Introduction

Symmetry plays a big role in physics. It often greatly simplifies the solution of a problem. If we have a perfectly round, concave bowl and throw a small marble into it, the task of following its subsequent trajectory would be quite complex. But if we are only interested in where it eventually comes to rest, the answer is clear: the only symmetric point, the centre (figure 1*a*). However, this does not always work. We could not solve the problem for a roulette wheel, because its structure explicitly breaks the symmetry. But suppose instead of a concave bowl, we have one shaped like the base of a wine bottle. Although this is still perfectly symmetric, the marble will not end up in the centre, where it would be sitting on a hump; it will come to rest somewhere on the circle of lowest points (figure 1*b*). This is an example of *spontaneous symmetry breaking*; the ground or lowest-energy state does not share the symmetry of the underlying physics. Instead, there is a whole family of ground states, the different points on the circle. The symmetry breaking is *spontaneous* in the sense that (unless we have extra information) we cannot predict which of these ground states will be chosen.

THE ROYAL SOCIETY PUBLISHING

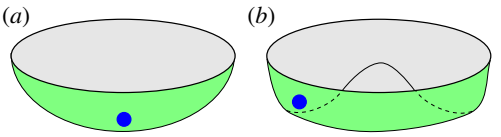

Figure 1. Final equilibrium point in a circular concave bowl (*a*) and a bowl with a central hump (*b*). (Online version in colour.)

Spontaneous symmetry breaking is ubiquitous in condensed matter physics. It often occurs when there is a phase transition between a high-temperature, symmetric phase and a low-temperature one in which the symmetry is spontaneously broken. The simplest example is freezing. If we have a round bowl of water sitting on a table, it looks the same from every direction; it has rotational symmetry. But when it freezes, the ice crystals form in specific orientations, breaking the symmetry.

Remarkably, the same phenomenon plays a crucial role in one part of the Standard Model of particle physics, the unified theory of weak and electromagnetic interactions first developed in the 1960s. In this paper, I describe the history of that development as I saw it from my vantage point as a member of the Theoretical Physics Group at Imperial College London, UK, led by Abdus Salam.

2. Particle physics after the Second World War

I have always believed that it was stroke of immense good fortune that I was able to join that group in 1959, less than 3 years after it was founded. Salam was a very talented physicist and a charismatic and inspiring leader, who attracted many distinguished visitors, such as Murray Gell-Mann and Steven Weinberg. It was a very exciting place to be, and a very exciting time in particle physics.

In quantum physics, as is well known, the distinction between particles and waves is lost; electrons show interference, and light comes in quantized photons. Both are described by *quantum fields*. The first quantum field theory to reach a mature form was *quantum electrodynamics* (QED), which describes the interactions of electrons and photons. In this theory, the interaction between two electrons is mediated by the exchange of one or more photons between them, as represented in figure 2.

QED had been around since the 1930s, and was very successful up to a point. The standard method of calculation, perturbation theory, involved a power-series expansion in powers of the *fine-structure constant*

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} \approx \frac{1}{137}, \tag{2.1}$$

where $-e$ is the charge on the electron, and $\hbar$ is Planck's constant divided by 2π. Lowest-order calculation, as represented by figure 2*a*, gave results in good agreement with experiment. (The photon is traditionally labelled γ and represented by a wiggly line, a relic of the idea that light is formed of waves. Conventionally, time goes upwards in these diagrams.) The trouble was that all the higher-order corrections, represented by diagrams such as those of figure 2*b*,*c*, gave infinite answers.

The solution to this problem, the concept of *renormalization*, was discovered in 1947 independently by Schwinger [1] and Feynman [2], and in fact earlier, in 1943, by Sin-Itiro Tomonaga, whose work in war-time Japan took some time to reach the West [3]. What these authors found was that it was possible to group all the infinities into corrections to the mass and charge of the electron. An electron is always surrounded by a cloud of 'virtual' photons and electron–positron pairs, whose presence changes the measured mass and charge. And if the results are re-expressed in terms of the actual measured mass and charge rather than the original 'bare' parameters, the infinities miraculously drop out.

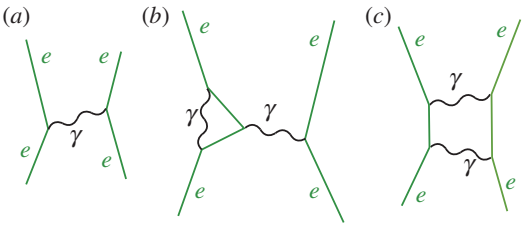

Figure 2. Electron–electron scattering: (*a*) lowest order, (*b*) and (*c*) higher order. (Online version in colour.)

Even today, this seems a slightly questionable procedure, but it works astonishingly well. The predicted results were verified by beautifully precise measurements of quantities such as the magnetic moment of the electron and the *Lamb shift* in hydrogen, the difference in energy between the $2p_{1/2}$ and $2s_{1/2}$ energy levels. QED rapidly became the most precisely verified theory in physics.

Following this success, the next natural goal was to find similarly successful theories of the other interactions or, even better, a unified theory of all of them—something for which we are still searching. Conventionally, we distinguish four types of interaction. Two of these are the familiar long-range forces of electromagnetism and gravity. The other two are short-range forces, negligible on scales much larger than an atomic nucleus. There is the strong nuclear force which binds together the protons and neutrons in nuclei, and the weak nuclear interaction responsible for radioactive beta decay, which also plays a key role in the energy-generation mechanism in the Sun. Each of these is characterized by a strength parameter or coupling constant analogous to the fine-structure constant. As things stood in the 1950s, it appeared that, for strong interactions, this constant must be of order one while for the weak it was around 10^{-10}. For the fourth interaction, gravity, it is of order 10^{-40}; on particle physics scales, gravity is totally negligible.

Initially, most interest centred on finding a theory of strong interactions, and some promising models were suggested. The problem was, however, that, with a coupling constant of order unity, perturbation theory is no use; successive terms in the series are all of the same order, so no one could actually do meaningful calculations. For this reason, quantum field theory fell out of fashion. When I was a student in the late 1950s, S-matrix theory, based on a study of the analytic properties of scattering amplitudes, was all the rage. Nevertheless, there were places, such as Imperial College—another was Harvard University in Cambridge, MA, USA—where the flag of field theory was kept flying.

3. Hadronic symmetries

Experimenters had been very busy too in the years after 1945. Observations of cosmic rays using cloud chambers, and later, at accelerators, of particle collisions in bubble chambers, revealed a whole host of new particles, especially hadrons, particles that have strong interactions, such as the proton and neutron. Within a few years, a 'zoo' of a hundred or more apparently elementary particles had emerged. One of the principal concerns of theorists was to try to bring some order into this chaos, by finding patterns and regularities, much as chemists had done in the previous century in establishing the periodic table.

A key role in this enterprise was played by the idea of approximate symmetries. The first of these, now called *isospin*, was well established, having already been proposed by Heisenberg [4]. Heisenberg noted that in many respects protons and neutrons are very similar; they have the same spin ($\frac{1}{2}$ in units of $\hbar$), almost the same mass, and essentially the same strong interactions. Heisenberg suggested that they might be regarded as two distinct states of a single entity, the *nucleon*; it would have two possible charge states, $N^+ = p$ and $N^0 = n$, analogous to the two spin states of an electron, $e_\uparrow$ and $e_\downarrow$. More importantly, Heisenberg suggested a *symmetry*, under

which one can 'rotate' one charge state into a combination of the two, just as one can rotate the spin of an electron. The symmetry is called 'isospin' not because of any physical connection to angular momentum, but because of a mathematical analogy with electron spin; both are described mathematically by the symmetry group SU(2). This proved to be a very prescient idea that was extremely useful for example in classifying the energy levels of light nuclei. It was later extended by Kemmer [5] to the other known strongly interacting particles, the pions (or π-mesons) which appear in three charge states (π^+, π^0, π^-).

Of course, this is not an *exact* symmetry. The proton and neutron do differ: the proton has a charge and the neutron does not. The symmetry is *broken* by the electromagnetic interactions. But because on the nuclear scale they are much weaker than the strong interactions, it is still useful. Today, we know that all hadrons are composed of quarks and antiquarks, and we understand isospin as a symmetry relating the two lightest quarks, up (u) and down (d), the constituents of nucleons and pions.

Another big step was taken in 1961. It was found that hadrons could be grouped in two-dimensional patterns of octets and decuplets, with triangular symmetry. This was explained, independently, by Gell-Mann [6] and Yuval Ne'eman [7], then a student of Salam's, by invoking a larger and more approximate SU(3) symmetry, which Gell-Mann called the 'eightfold way'. This is now understood as a symmetry of the *three* lightest quarks, u, d and s (strange).

4. Gauge theories

QED is a *gauge theory*. This means it has a special kind of *local* symmetry. In ordinary quantum mechanics, the physics is unchanged by a global change of phase of the wave function ψ:

$$\psi(t, \mathbf{r}) \to \psi(t, \mathbf{r}) e^{i\alpha}, \tag{4.1}$$

where α is an arbitrary constant. Mathematically, this is a U(1) symmetry. In quantum *field theory*, however, the physics is unchanged by a more general space–time-dependent transformation:

$$\psi(t, \mathbf{r}) \to \psi(t, \mathbf{r}) e^{i\alpha(t,\mathbf{r})}, \tag{4.2}$$

provided that a compensating transformation is applied to the electromagnetic potentials.

This idea of replacing a global by a local symmetry, often called the 'gauge principle', provides a sort of rationale for the existence of the electromagnetic field; without that field, the extension is not possible. Moreover, it was also seen as offering an explanation for the fact that the quantum of the electromagnetic field, the photon, is massless, in the sense of having zero rest-mass; in vacuum it cannot ever be at rest, and its minimum energy is zero. This is because the gauge principle does not work if we add a mass term. That yields a theory that cannot be renormalized, so it is inconsistent in the sense of giving infinite answers to physical questions.

It seemed natural to ask if this gauge principle could be extended. Salam was convinced from an early stage that a unified theory of all interactions should be a gauge theory. The first proposal of this kind was made by Yang & Mills [8] in 1954 and in fact independently in the same year by Ronald Shaw, then a student of Salam's, though he never published it except as a PhD thesis [9]. This Yang–Mills theory was based on applying the gauge principle to the isospin SU(2) group, and was intended as a theory of strong interactions. But the fact that isospin is not an exact symmetry posed a problem, because adding symmetry-breaking terms seemed to destroy the nice properties of gauge theories. In the end, it did not turn out to be the correct theory of strong interactions; that is a quite different gauge theory called *quantum chromodynamics* (*QCD*), which I will not discuss in detail here. However, the Yang–Mills theory is the basis on which all subsequent gauge theories have been founded.

Because of the difficulty, mentioned earlier, of calculating with a strongly interacting theory, interest began at least to some extent to shift towards the weak interactions, especially after the

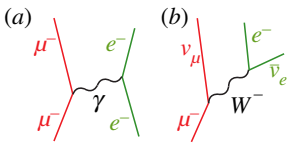

Figure 3. (*a*) Muon–electron scattering and (*b*) muon decay. (Online version in colour.)

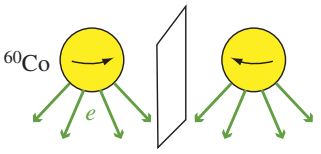

Figure 4. Decay of ^{60}Co, viewed directly and in a mirror image. (Online version in colour.)

discovery that they could be understood as proceeding through the exchange of particles of spin 1 (in units of $\hbar$), such as the photon [10,11]. In this case, two charged particles, called W^+ and W^-, were required. So people began to ask, could this be a gauge theory, or even could there be a *unified* theory of weak and electromagnetic interactions, with some sort of symmetry between the three 'gauge bosons', the W^+, W^- and γ?

It is worth comparing examples of the two kinds of interactions. Figure 3*a* shows a typical electromagnetic interaction, the scattering of an electron and a muon (a heavier version of the electron), proceeding via photon exchange. Figure 3*b*, on the other hand, represents a typical weak process, the decay of a muon into an electron and a pair of neutrinos, proceeding via the intermediary of a W^-. Clearly, these two diagrams are quite similar. But there are some very important differences. Most significantly, the long range of the electromagnetic force means that the photon must have zero rest-mass, as expected for a gauge boson. But correspondingly, the very short range of the weak interactions means that the W^+ and W^- must have very large masses, of the order of 100 times the mass of a proton. Moreover, the electromagnetic interactions conserve parity, meaning that they look the same if viewed through a mirror. But the weak interactions do not. It was discovered in 1957 that they violate parity conservation; spinning ^{60}Co nuclei, when they decay, emit electrons preferentially in the direction opposite to the spin, a correlation that would be reversed if we looked through a mirror [12] (figure 4). So clearly if there is some kind of symmetry between the three gauge bosons, it must be *broken*.

The first proposal of a gauge theory of weak interactions came in 1957 from Schwinger [13], who also suggested there might be a unified theory based on the SU(2) group involving all three gauge bosons. A big step forward was taken by Glashow in 1961 [14]. To solve the parity problem, he proposed an extended theory, based on a larger symmetry group SU(2)×U(1) and incorporating a fourth gauge boson Z^0, and showed that, by a mixing mechanism between the two neutral gauge bosons, it was possible to arrange that one (the photon) would have parity-conserving interactions, whereas the others, such as the $W^\pm$, did not.

Salam & Ward (long-time collaborator of Salam) [15], apparently unaware of Glashow's work, wrote down a very similar model, also based on SU(2) × U(1), 3 years later.

But in all these models, the symmetry breaking, giving the $W^\pm$ large masses, had to be inserted by hand, and models with spin-1 bosons with explicit masses were well known to be non-renormalizable, in other words inconsistent.

The big question that began to be asked was whether this could be achieved by *spontaneous* symmetry breaking, where the underlying symmetry remains, but is broken by its particular realization. It was first suggested by Nambu [16,17] that particle masses might be generated by such a process, by analogy with what was known to happen in superconductivity. There was, however, a major difficulty.

5. Nambu–Goldstone bosons

The problem was that in many cases spontaneous breaking of a continuous symmetry leads to the appearance of unwanted massless scalar (spin-0) bosons [17,18]: 'unwanted' because no one had seen any such bosons, and if they existed and had any reasonable strength of interaction they ought to have been easy to see.

The simplest model to illustrate this is the Goldstone model, which involves a complex scalar field $\phi(t, \mathbf{r})$ with a self-interaction described by the potential function

$$\mathcal{V} = \tfrac{1}{2}\lambda(\phi^*\phi - \tfrac{1}{2}\eta^2)^2, \tag{5.1}$$

where λ and η are positive constants. This is the *sombrero potential* (figure 5).

Like the base of a wine bottle discussed earlier, this potential has a maximum at $\phi = 0$, and minima around a circle. So the ground state, in this case the vacuum state $|0\rangle$, breaks the symmetry. In fact, we have a *degenerate* vacuum state; we can have

$$\langle 0|\phi|0\rangle = \frac{\eta}{\sqrt{2}}e^{i\alpha}, \tag{5.2}$$

for any value of α.

Suppose we choose one particular value, say $\alpha = 0$, thus making the expectation value real, and write ϕ as the vacuum expectation value plus real and imaginary parts:

$$\phi = \frac{1}{\sqrt{2}}(\eta + \varphi_1 + i\varphi_2). \tag{5.3}$$

Then, substituting into $\mathcal{V}$, we find

$$\mathcal{V} = \tfrac{1}{2}m_1^2\varphi_1^2 + \text{cubic and quartic terms}, \tag{5.4}$$

where $m_1^2 = \lambda\eta^2$. The mass of the φ_1 boson representing radial oscillations is determined by the curvature of the potential in that direction. But in the transverse φ_2 direction, the potential is flat, and the corresponding boson is massless. This is the *Nambu–Goldstone boson*. The corresponding waves represent space-dependent oscillations in the direction of symmetry breaking.

If this phenomenon were inevitable, it seemed that spontaneous symmetry breaking could be ruled out, because nobody had seen any massless spin-0 bosons. There were known counterexamples in condensed matter physics, in particular superconductivity [16,19]. But these were not well understood; there was considerable debate among the experts as to whether that was really an example of spontaneous symmetry breaking, and it was generally believed that the Nambu–Goldstone bosons were unavoidable in relativistic theories. This became known as the *Goldstone theorem*.

Naturally, there was a lot of discussion of this issue among those trying to develop gauge theories. When Steven Weinberg came to Imperial College on sabbatical in 1961/1962, he and Salam spent a lot of time discussing this issue, and eventually developed what seemed like a watertight proof, which they published together with Goldstone [20].

For the sake of those who want to know the details, let me summarize the argument. (Others may skip this paragraph.) It relied on the well-known fact (called Noether's theorem after its discoverer, Emmy Noether) that a continuous symmetry is always associated with a conservation law; for example, translational symmetry is associated with conservation of momentum. The conservation law is expressed through a local continuity equation of the form

$$\partial_\mu j^\mu \equiv \frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{j} = 0, \tag{5.5}$$

where $j^0 \equiv \rho$ is the density of the conserved quantity and $\mathbf{j} \equiv (j^1, j^2, j^3)$ the flux density. This normally leads to a global conservation law

$$\frac{dQ}{dt} = 0, \quad \text{where} \quad Q(t) = \int \rho(t, \mathbf{r})d^3\mathbf{r}. \tag{5.6}$$

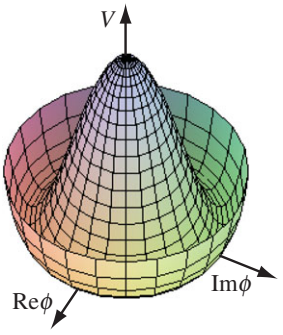

Figure 5. The sombrero potential. (Online version in colour.)

The conserved quantity Q generates the symmetry in the sense that, under an infinitesimal symmetry transformation, the change in any field ϕ is given by the commutator

$$\delta\phi(t,\mathbf{r}) = \mathrm{i}\epsilon[\phi(t,\mathbf{r}), Q(t)], \tag{5.7}$$

where ϵ is an infinitesimal parameter. The fact that the symmetry is spontaneously broken means that there is some field ϕ whose vacuum expectation value $\langle 0|\phi|0\rangle$ is not invariant under the transformation, i.e. $\langle 0|[\phi, Q]|0\rangle \neq 0$. However, if Q is time-independent, this can happen only if there are possible intermediate states with zero energy in the limit of zero momentum, in other words massless particle states.

The development seemed to have reached an impasse. Spontaneous symmetry breaking required the existence of massless spin-0 particles, which ought to have been easy to see. Because none had been found, that appeared to rule the idea out. Weinberg commented on this disappointing conclusion 'Nothing will come of nothing; speak again!', a quotation from *King Lear*. Fortunately, however, we were eventually able to speak again.

6. The mass generation mechanism

In 1964, a young American postdoctoral fellow arrived at Imperial College, Gerald Guralnik, who had been a student of Walter Gilbert, who in turn was a student of Salam. I was very interested to find that he had already been working on this problem and indeed published some ideas about it [21,22]. We began working on it, together with another American visitor, Carl Richard Hagen. We, and of course others, found the solution. It turns out that gauge theories are different. In a remarkable way, the masslessness of the would-be Nambu–Goldstone boson and of the gauge boson apparently 'cancel out', creating a massive gauge boson.

This was presented in papers by three independent groups, all published in the summer and autumn of 1964: first by Englert & Brout [23] from Brussels, Belgium, then Higgs [24,25] from Edinburgh, UK, and finally Guralnik *et al*. [26] from Imperial College.

All three, starting from very different viewpoints, proposed essentially the same model for spontaneous symmetry breaking in the simplest U(1) gauge theory, a broken version of electrodynamics of spin-0 charged particles. It involves introducing a new complex scalar field ϕ. To be specific, again for those who want the details, it is described by the Lagrangian density

$$\mathcal{L} = D_\mu\phi^* D^\mu\phi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - \mathcal{V}(\phi), \tag{6.1}$$

where

$$D_\mu\phi = \partial_\mu\phi + \mathrm{i}eA_\mu\phi, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{6.2}$$

and $\mathcal{V}$ is the same sombrero potential (5.4) as in the Goldstone model. As before, the field acquires a non-zero average value, breaking the symmetry spontaneously. We can again make the same expansion of ϕ as in (5.3), and it is convenient also to define a new field B_μ by

$$B_\mu = A_\mu + \frac{1}{e\eta}\partial_\mu \varphi_2. \tag{6.3}$$

The relation between A_μ and B_μ is effectively a gauge transformation, so $F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$. The radial φ_1 mode acquires a mass as before from the curvature of the potential; this is the Higgs boson. But, the kinetic term for φ_2 turns into a mass term for B_μ. We find

$$\mathcal{L} = \tfrac{1}{2}\partial_\mu\varphi_1\partial^\mu\varphi_1 - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - \tfrac{1}{2}\lambda\eta^2\varphi_1^2 + \tfrac{1}{2}e^2\eta^2 B_\mu B^\mu + \text{cubic and quartic terms}. \tag{6.4}$$

Miraculously, the massless fields seem to have disappeared, leaving only a massive scalar (the Higgs boson) and a massive gauge (vector) boson.

7. Escaping the Goldstone theorem

How then did we manage to avoid the conclusions of the Goldstone theorem? This is a very subtle question, which I try to answer in this section. (Those who are not interested in these details can safely skip to the next.)

The answer lies in a well-known peculiarity of the electromagnetic field (and other gauge fields). Gauge invariance means that we can make a transformation of the form (4.2), together with the matching transformation of the electromagnetic potentials, without changing anything physical. Thus, the same physical state can be represented by many different solutions of the field equations related to each other in this way. Correspondingly, there may be oscillations in the field variables that look like physical particles, but actually represent only oscillations in the value of the function α. These are not real physical particles but only 'gauge modes', space–time-dependent gauge transformations. For many purposes, this is awkward, and it is convenient to impose a specific gauge condition, a rule for choosing one specific representation for each physical state.

One of the simplest is the so-called Coulomb or radiation gauge condition, the requirement that the spatial part of the potential field A_μ be divergenceless:

$$\partial_k A^k \equiv \nabla\cdot\mathbf{A} = 0. \tag{7.1}$$

If $B^\mu = 0$ then by (6.3) this implies that φ_2 must satisfy the Laplace equation, $\nabla^2\varphi_2 = 0$, which, with appropriate boundary conditions, means that φ_2 itself vanishes (or is at most a constant). But imposing the Coulomb-gauge condition (7.1) requires that we choose a specific rest frame. By imposing this condition we remove the unwanted gauge modes, but we also lose the explicit relativistic invariance of the theory under Lorentz transformations. The physical results are of course still independent of the choice of reference frame, but if we go to a relatively moving frame we also have to make a gauge transformation to impose the Coulomb-gauge condition in the new frame.

Now, one of the assumptions made in proving the Goldstone theorem [20] was that the theory is explicitly Lorentz invariant. This is the assumption that fails in the case of Coulomb-gauge electrodynamics. The time independence of the 'charge' operator Q of (5.6) only follows from the continuity equation (5.5) if one can drop a surface integral term at infinity. This is legitimate in a manifestly Lorentz-invariant theory, because what we are really interested in are commutators, and in such a theory the commutators always vanish outside the light cone. However, it is not legitimate in Coulomb-gauge electrodynamics, where commutators fall off quite slowly at large distances. In fact, as was shown explicitly in our paper [26], the operator Q, in this case, not only fails to be time-independent, but actually does not exist at all as a well-defined operator. This is now recognized as a characteristic feature of spontaneous symmetry breaking.

There is an alternative approach that does preserve relativistic Lorentz invariance. That is, to choose instead of (7.1) the Lorentz gauge condition, namely

$$\partial_\mu A^\mu = 0, \tag{7.2}$$

which does not require a choice of reference frame. But, in this case, some gauge freedom remains; here, if $B^\mu = 0$, all we can say about φ_2 is that it satisfies the wave equation, $\partial_\mu \partial^\mu \varphi_2 = 0$, which, being a hyperbolic equation, does have non-trivial solutions. In this manifestly Lorentz-invariant formalism, the Goldstone theorem does apply. There are Nambu–Goldstone bosons, and because the field φ_2 satisfies the wave equation they are massless. But they are not physical, they are pure gauge modes.

There is one further subtlety. We have spontaneous symmetry breaking in a gauge theory, but it is not, in fact, the local gauge transformations that are spontaneously broken. Indeed, in 1975, Elitzur [27] showed that spontaneous breaking of a local symmetry is logically impossible. In the Coulomb-gauge formalism, our gauge condition has eliminated the local gauge invariance; it is broken explicitly, not spontaneously. But there remains a global symmetry. We chose ϕ to be real, or equivalently $\varphi_2 = 0$, but we could have chosen any other fixed phase instead. (This corresponds to the fact that φ_2 may not vanish but be a non-zero constant.) It is this remaining global symmetry that is actually spontaneously broken. Moreover, it should be emphasized that, unlike for example the distinct degenerate ground states of a freezing liquid, our distinct vacuum states are all physically equivalent. This is because for the phase of ϕ, unlike the orientation of the crystals, there is no fixed external reference standard. In our case, there is no true physical degeneracy of the vacuum.

8. Electroweak unification

The papers from the three groups I have described attracted hardly any attention at the time, except for some scepticism. By the end of 1964, we knew of Glashow's (or Salam and Ward's) SU(2) × U(1) model, and we knew of the mass-generating mechanism based on spontaneous symmetry breaking. So it is perhaps rather surprising that it still took three more years for anyone to think of putting the two together. This may have been partly because many of us were still thinking more of strong interactions.

I did some further work early in 1967 on how the mechanism could actually be applied to more realistic gauge-theory models with larger symmetry groups, in particular on how the symmetry-breaking pattern determines the numbers of massive and massless bosons [28]. I had several discussions with Salam which I think served to revive his interest in the subject. Eventually, a unified theory of weak and electromagnetic interactions of leptons (particles such as the electron and muon that do not have strong interactions) was proposed by Weinberg [29]. Salam presented essentially the same model in lectures he gave at Imperial College in the autumn of 1967 (though I was not present, because I was then in the USA). He did not publish this work until the following year [30], probably, in large measure, because he was very occupied with developments at his International Centre for Theoretical Physics in Trieste, Italy. He called this the *electroweak model*.

Both Salam and Weinberg speculated that their theory was renormalizable, but they were unable to prove it. It was proved in 1971 by a young student, Gerard 't Hooft [31], in a real *tour de force* using methods developed by his supervisor, Martinus Veltman, especially the computer algebra program *Schoonship*.

In 1973, the key prediction of the theory, the existence of 'neutral current' interactions, those mediated by Z^0, was established at CERN [32]. This led to the award of the Nobel Prize to Glashow, Salam and Weinberg in 1979, but Ward was left out, even though he was a collaborator on almost all of Salam's papers on the subject, perhaps because of the 'rule of three' that limits the number of recipients to no more than three.

In 1983, the W and Z particles were finally discovered at CERN [33,34]. This led the following year to Nobel Prizes for the experiments to Carlo Rubbia and Simon van der Meer.

't Hooft and Veltman gained their Nobel Prizes much later, in 1999.

Also during the 1970s and 1980s, in parallel with these developments, a theory of strong interactions, quantum chromodynamics (QCD), was developed. This was a gauge theory based on an SU(3) symmetry, though with no relation to the eightfold way. It does not involve any spontaneously broken symmetry—the SU(3) is exact—but solves the problem of short range by a completely different mechanism. So by then we had a *Standard Model* of particle physics, based on the symmetry group SU(3) × SU(2) × U(1). Over the past few decades, this model has been tested in all kinds of ways and become more and more established.

9. The Higgs boson

In 1964 or 1967, the Higgs boson had been a rather minor feature; the important thing was the *mechanism* that gave masses to the gauge bosons while escaping the Goldstone theorem. But after 1983, when all the other elements of the Standard Model had been seen (except the top quark, which completes the family of six), the Higgs boson started to acquire a key importance, as the last remaining piece of the standard-model jigsaw. The Standard Model worked so well that it (or something else doing the same job) more or less *had* to be present.

Finding the Higgs boson was of course one of main objectives of the *Large Hadron Collider* (LHC) at CERN, and of the two dedicated teams that designed, built and operated the great ATLAS and CMS detectors. The search was triumphantly successful in 2012 [35,36], leading to Nobel Prizes for Englert and Higgs in 2013.

The main purpose of the Higgs boson was to give mass to gauge bosons, but it also gives mass to all other particles it interacts with. The mechanism is very similar to refraction. The non-zero vacuum value of the Higgs field corresponds in particle terms to a uniform sea or *condensate* of Higgs bosons. A similar effect is produced when a photon enters a plasma, containing free electrons. The photon is slowed by interaction with the electrons, though without losing energy. It acquires an effective mass m_γ related to the *plasma frequency* $\omega_{\rm pl}$ by

$$m_\gamma = \frac{\hbar\omega_{\rm pl}}{c^2}, \quad \text{with} \quad \omega_{\rm pl}^2 = \frac{e^2 n_{\rm e}}{\epsilon_0 m_e}, \tag{9.1}$$

where $n_{\rm e}$ is the electron density. Here, the electrons play an analogous role to the Higgs bosons.

It is quite often said that the Higgs is responsible for all the mass in the Universe. But this is not true. It does give mass to all the known elementary particles except the photon, which remains massless, and perhaps the neutrinos, which do have small, but non-zero masses of unknown origin. But it does not contribute most of the mass of protons and neutrons, which comes instead from the gluons (the gauge bosons of QCD) that hold them together.

10. Conclusion

Our Standard Model is now in one sense complete. But, it is certainly far from being the last word. It is a wonderfully successful theory that agrees very well with almost all the experimental data. But it must be admitted that it is also a mess. It has something like 20 arbitrary parameters, such as mass ratios and coupling strengths, that we cannot predict. Many of these are apparently quite critical for our very existence, so one theory is that there are many universes in which these parameters take all kinds of different values, but creatures like us will only appear in a very restricted set in which the parameters happen to have values favourable to us. To me however that *anthropic* theory seems like a counsel of despair.

Moreover, there are still a lot of things the Standard Model does not explain: what is the dark matter and the 'dark energy' in the Universe?; why did it apparently undergo a rapid period of inflation?; why do the neutrinos have these tiny masses?; why are there, as there seem to be, three generations of otherwise similar particles with wildly differing masses? In addition, the standard model is not truly unified, because its symmetry is a product of three distinct symmetries, each of which has its own coupling strength. And finally, of course, the Standard Model does not include gravity, and efforts to bring the two together have shown that there are major obstacles. Possibly,

for this, we need string theory or its modern incarnation, M-theory, or loop quantum gravity, or who knows what?

Acknowledgements. I thank the organizers for inviting me to give a talk at the 'Before, behind and beyond the discovery of the Higgs boson' meeting, on which this article is based. I would also like to acknowledge the huge debt I owe to my mentor and inspiration, Abdus Salam, whose premature death in 1996 was a tragic loss, as well as to my erstwhile collaborator and lifelong friend Gerald Guralnik, who sadly died of a sudden heart attack in April this year.

References

1. Schwinger JS. 1948 On quantum electrodynamics and the magnetic moment of the electron. *Phys. Rev.* **73**, 416–417. (doi:10.1103/PhysRev.73.416)
2. Feynman RP. 1948 Relativistic cutoff for quantum electrodynamics. *Phys. Rev.* **74**, 1430–1438. (doi:10.1103/PhysRev.74.1430)
3. Tomonaga S-I. 1946 On a relativistically invariant formulation of the quantum theory of wave fields. *Prog. Theor. Phys.* **1**, 27–42. (doi:10.1143/PTP.1.27)
4. Heisenberg W. 1932 Über den Bau der Atomkerne. I. *Z. Phys.* **77**, 1–11. (doi:10.1007/BF01342433)
5. Kemmer N. 1938 The charge-dependence of nuclear forces. *Proc. Camb. Phil. Soc.* **34**, 354–364. (doi:10.1017/S0305004100020296)
6. Gell-Mann M. 1961 The eightfold way: a theory of strong interaction symmetry. California Institute of Technology Synchrotron Laboratory Report, CTSL-20. See http://www.osti.gov/scitech/servlets/purl/4008239.
7. Ne'eman Y. 1961 Derivation of strong interactions from a gauge invariance. *Nucl. Phys.* **26**, 222–229. (doi:10.1016/0029-5582(61)90134-1)
8. Yang CN, Mills RL. 1954 Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.* **96**, 191–195. (doi:10.1103/PhysRev.96.191)
9. Shaw R. 1955 Invariance under general isotopic gauge transformations, part II, chapter III. PhD thesis, University of Cambridge, UK.
10. Feynman RP, Gell-Mann M. 1958 Theory of the Fermi interaction. *Phys. Rev.* **109**, 193–197. (doi:10.1103/PhysRev.109.193)
11. Sudarshan ECG, Marshak RE. 1958 Chirality invariance and the universal Fermi interaction. *Phys. Rev.* **109**, 1860–1862. (doi:10.1103/PhysRev.109.1860.2)
12. Wu CS, Ambler E, Hayward RW, Hoppes DD, Hudson RP. 1957 Experimental test of parity conservation in beta decay. *Phys. Rev.* **105**, 1413–1415. (doi:10.1103/PhysRev.105.1413)
13. Schwinger JS. 1957 A theory of the fundamental interactions. *Ann. Phys.* **2**, 407–434. (doi:10.1016/0003-4916(57)90015-5)
14. Glashow SL. 1961 Partial symmetries of weak interactions. *Nucl. Phys.* **22**, 579–588. (doi:10.1016/0029-5582(61)90469-2)
15. Salam A, Ward JC. 1964 Electromagnetic and weak interactions. *Phys. Lett.* **13**, 168–171. (doi:10.1016/0031-9163(64)90711-5)
16. Nambu Y. 1960 Quasi-particles and gauge invariance in the theory of superconductivity. *Phys. Rev.* **117**, 648–663. (doi:10.1103/PhysRev.117.648)
17. Nambu Y, Jona-Lasinio G. 1961 Dynamical model of elementary particles based on an analogy with superconductivity. I. *Phys. Rev.* **122**, 345–358. (doi:10.1103/PhysRev.122.345)
18. Goldstone J. 1961 Field theories with superconductor solutions. *Nuovo Cim.* **19**, 154–164. (doi:10.1007/BF02812722)
19. Anderson PW. 1963 Plasmons, gauge invariance, and mass. *Phys. Rev.* **130**, 439–442. (doi:10.1103/PhysRev.130.439)
20. Goldstone J, Salam A, Weinberg S. 1962 Broken symmetries. *Phys. Rev.* **127**, 965–970. (doi:10.1103/PhysRev.127.965)
21. Guralnik GS. 1964 Photon as a symmetry-breaking solution to field theory. I. *Phys. Rev.* **136**, B1404–B1416. (doi:10.1103/PhysRev.136.B1404)
22. Guralnik GS. 1964 Photon as a symmetry-breaking solution to field theory. II. *Phys. Rev.* **136**, B1417–B1422. (doi:10.1103/PhysRev.136.B1417)
23. Englert F, Brout R. 1964 Broken symmetry and the mass of gauge vector bosons. *Phys. Rev. Lett.* **13**, 321–323. (doi:10.1103/PhysRevLett.13.321)

24. Higgs PW. 1964 Broken symmetries, massless particles and gauge fields. *Phys. Lett.* **12**, 132–133. (doi:10.1016/0031-9163(64)91136-9)
25. Higgs PW. 1964 Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.* **13**, 508–509. (doi:10.1103/PhysRevLett.13.508)
26. Guralnik GS, Hagen CR, Kibble TWB. 1964 Global conservation laws and massless particles. *Phys. Rev. Lett.* **13**, 585–587. (doi:10.1103/PhysRevLett.13.585)
27. Elitzur S. 1975 Impossibility of spontaneously breaking local symmetries. *Phys. Rev. D* **12**, 3978–3982. (doi:10.1103/PhysRevD.12.3978)
28. Kibble TWB. 1967 Symmetry breaking in non-Abelian gauge theories. *Phys. Rev.* **155**, 1554–1561. (doi:10.1103/PhysRev.155.1554)
29. Weinberg S. 1967 A model of leptons. *Phys. Rev. Lett.* **19**, 1264–1266. (doi:10.1103/PhysRevLett.19.1264)
30. Salam A. 1968 Weak and electromagnetic interactions. In *Proc. of the 8th Nobel Symposium on Elementary particle theory, relativistic groups and analyticity, Lerum, Sweden, 19–25 May 1968* (ed. N Svartholm), pp. 367–377. Stockholm, Sweden: Almqvist & Wiksell.
31. 't Hooft G. 1971 Renormalizable Lagrangians for massive Yang–Mills fields. *Nucl. Phys. B* **35**, 167–188. (doi:10.1016/0550-3213(71)90139-8)
32. Hasert FJ *et al.* 1973 Observation of neutrino like interactions without muon or electron in the Gargamelle neutrino experiment. *Phys. Lett. B* **46**, 138–140. (doi:10.1016/0370-2693(73)90499-1)
33. Arnison G *et al.* 1983 Experimental observation of isolated large transverse energy electrons with associated missing energy at $s^{1/2} = 540$ GeV. *Phys. Lett. B* **122**, 103–116. (doi:10.1016/0370-2693(83)91177-2)
34. Banner M *et al.* 1983 Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN anti-p–p collider. *Phys. Lett. B* **122**, 476–485. (doi:10.1016/0370-2693(83)91605-2)
35. Aad G *et al.* 2012 Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B* **716**, 1–29. (doi:10.1016/j.physletb.2012.08.020)
36. Chatrchyan S *et al.* 2012 Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B* **716**, 30–61. (doi:10.1016/j.physletb.2012.08.021)