SPIN-2001/03

THE DISCOVERY OF THE RENORMALIZABILITY OF NON-ABELIAN GAUGE THEORIES

Gerard 't Hooft

Institute for Theoretical Physics
University of Utrecht, Princetonplein 5
3584 CC Utrecht, the Netherlands

and

Spinoza Institute
Postbox 80.195
3508 TD Utrecht, the Netherlands

e-mail: g.thooft@phys.uu.nl
internet: http://www.phys.uu.nl/~thooft/

ABSTRACT

A personal account is presented of the sequence of observations and discoveries that led to the discovery that non-Abelian gauge theories without anomalies are renormalizable.

1. INTRODUCTION.

Before 1970, the prevailing view of particle physicists on the elementary building blocks of matter was very different from today's. Although several successes of quantum field theory were well-established, and recognized as such, the general feeling was nevertheless that this was not the way to really understand what was going on. Quantum field theory was plagued by infinities, and it was therefore considered ugly. It seemed to be abundantly clear that the real world should be described by a more efficient scheme, without the need of perturbative approximations and without infinities that had to be "swept under the rug". [1,2]

Only very few physicists seriously studied local quantized fields that required renormalization, but among them was my then thesis advisor, M. Veltman. A 'must' for his students was to read a paper by C.N. Yang and R.L. Mills [3] of 1954. The paper made a daring assumption: a field was introduced with the property that particles with certain quantum numbers such as isospin, traversing this field, would undergo transitions in such a way that these quantum numbers change. This would promote isospin from a global symmetry to a local symmetry.

Now, 30 years later, this paper is famous, but at that time its relevance for elementary particle physics was far from clear. The theory appeared to require an internal degeneracy of the particle spectrum of a kind not seen in the experiments, and a massless particle with spin one that had to carry a charge sensitive to its own field, unlike the one and only known massless spin one particle, the photon.

If the Yang-Mills paper was nevertheless taken seriously at all, it was because it clearly showed an extraordinary elegance. It showed features closely resembling two well-established forces of Nature: electro-magnetism and gravity. Some investigators [4,5] took it as a toy model for gravity, others, such as Veltman [6], were intrigued by the universal strength of the Yang-Mills force. This feature does resemble the observed universality in the weak force.

The algebraic structure of the weak force was well-established in the '60s. It was known that this force can be described as if mediated by a very heavy bosonic particle [7]. The space-time point where this particle is created must be very close to the point where it is annihilated, and at each of these two space-time points, a fermion such as an electron transmutes into its partner, an electron-neutrino, or *vice versa.* Alternatively, at such a point a proton could transmute into a neutron, a lambda or *vice versa*, or, in a language already becoming familiar, *up* quarks would be interchanged by *down* or *strange* quarks. The exchanged particle has the space-time structure of a spin-one particle. Veltman concluded that its interactions had to be very similar to the Yang-Mills interaction; however, the exchanged particle must have a considerable amount of mass, whereas the Yang-Mills photon is massless.

At first sight, this departure from the pure Yang-Mills theory did not appear to be disastrous. It is possible to add a *mass term* into the Yang-Mills equations [4,6,8], and this way a theory was obtained that seemed to reproduce many of the features observed in the weak interactions. It was this theory that was to be investigated by the quantum field adherents.

2. FEYNMAN RULES AND CUTTING RELATIONS.

Under many circumstances, perturbation theory is a very powerful tool. A condition is that the coupling strengths of the theory are rather small. In that case, the successive corrections calculated for a scattering amplitude rapidly converge towards an accurately calculable result. The increasingly complicated expressions that need to be calculated can be conveniently described in terms of diagrams, the so-called Feynman diagrams. All scattering amplitudes together form a matrix in the mathematical sense, the *scattering matrix*. An amplitude is an element of this matrix, and it can be represented by the set of Feynman diagrams that need to be calculated to obtain its value.

In order for the theory to make sense, the scattering matrix is required to possess certain properties. One of these properties refers to the amplitudes associated to the probability that a particular event takes place, and in addition the amplitude for particles not to interact at all: the total sum of all probabilities must add up to one, and of course each of these probabilities must be non-negative. It turns out that this requirement implies that the scattering matrix must be unitary.

Furthermore, one insists that the scattering matrix can be represented in such a way that the interactions are seen to be well-ordered in time, so as to establish a certain sense of *causality*. Causality implies that the scattering matrix must obey certain relations, called *dispersion relations*. The simplest of these relations gives us restrictions in how wavelengths depend on frequency when particle waves go through matter: no information here is allowed to go faster than the speed of light.

For the renormalizable field theories known at the time, which were quantum electrodynamics and theories describing interactions among spinless fundamental particles, it could be derived that the requirements mentioned above are indeed obeyed if one uses the Feynman rules correctly. Of course, it would have been surprising if this were not so, since the Feynman rules follow from a perfectly consistent elaboration of these quantum field theories. One has to remember, however, that these theories also generate *infinities*, which had to be circumvented by renormalization. At this point, the unitarity relations and the dispersion relations became a tool; the only acceptable subtraction procedures are the ones that respect these relations. It turns out that this requirement is more than sufficient to determine the renormalization counter terms, but since they *over*determine these coefficients, there is the danger that they clash. Indeed, examples of such clashes are known; they are called 'anomalies' [9]. Theories with only scalar particles and particles with spin $^1/_2$ were known to be renormalizable without any such difficulties. There was only one renormalizable vector theory known: Quantum Electrodynamics (QED), the theory of the quantized electromagnetic field. Here, the primary quantum particles are the photons, particles with spin one and mass zero.

Now, the Yang-Mills theory that appeared to be most appropriate for describing the weak interaction had an explicitly added mass term in it. [8] With that, the Feynman rules yielded a *propagator* for this vector field that had the following form:

$$P_{\mu\nu}(k) = \frac{\delta_{\mu\nu} + k_\mu k_\nu / M^2}{k^2 + M^2 - i\varepsilon} \ , \tag{2.1}$$

where M is the mass of the vector particle. The $k_\mu k_\nu$ term appeared to be disastrous, since it gives rise to bad divergences as k_μ tends to infinity, divergences closely related to the fact that the limit $M \to 0$ appears to be singular. Now, in electrodynamics, the $k_\mu k_\nu$ terms are harmless, because the propagator is coupled to a conserved current J_μ, which, in k-space, has $k_\mu J_\mu(k) = 0$. Conservation of the current is guaranteed by gauge-invariance. However, in the weak interactions, an intermediate vector boson couples to a non conserved current, since under its action, a particles transmutes into a different one, with different mass, and furthermore, the field couples to itself. Anyway, gauge invariance is broken by the mass term. How should one proceed here?

Why was the $k_\mu k_\nu$ term needed anyway? To understand this, we have to analyze the mathematical structure of quantum field theories. A consistent theory must yield a scattering matrix that is unitary and obeys certain dispersion relations, as was explained above. These properties can indeed be derived from the Feynman rules. This should not be surprising, since the Feynman rules had been derived from consistent theories, but how can we see this directly? Veltman had studied the papers by Bogolyubov et al [10] on this, and eventually came with his own formulation, which is very transparent. It goes as follows.

3. CUTTING RULES. [11]

In scalar theories, a propagator in momentum space may be written as

$$\Delta(k) = -\int_0^\infty \frac{\varrho(m)\mathrm{d}m}{i(k^2+m^2-i\varepsilon)} \ , \tag{3.1}$$

where $\varrho(m)$, which is only non vanishing for non negative m, represents a distribution of possible mass values. It may consist of one or more Dirac delta functions, when a particle is stable, but it may also describe a resonance curve.

In coordinate space, the propagator depends on two space-time points, $x^{(1)}$ and $x^{(2)}$, and it is written as

$$\Delta(x^{(2)} - x^{(1)}) = \frac{1}{\sqrt{(2\pi)^4}} \int \mathrm{d}^4k \Delta(k) e^{ik\cdot(x^{(2)}-x^{(1)})} \ . \tag{3.2}$$

In addition, we introduce the *on-shell propagator*, $\Delta^+(k)$ as follows:

$$\Delta^+(k) = 2\pi \int_0^\infty \varrho(m)\mathrm{d}m\, \theta(k_0)\delta(k^2+m^2) \, , \tag{3.3}$$

and similarly,

$$\Delta^-(k) = 2\pi \int_0^\infty \varrho(m)\mathrm{d}m\, \theta(-k_0)\delta(k^2+m^2) \, . \tag{3.4}$$

We see that Δ^+ describes a particle with positive energy, and Δ^- one with negative energy. The transition towards coordinate space is now as in Eq. (3.2).

from elementary algebra it now follows that

$$\begin{aligned} \Delta(x^{(2)} - x^{(1)}) &= \Delta^+(x^{(2)} - x^{(1)}) \qquad \text{if } t^2 \ge t^1 \ , \\ \text{and } \ \Delta(x^{(2)} - x^{(1)}) &= \Delta^-(x^{(2)} - x^{(1)}) \qquad \text{if } t^2 \le t^1 \ , \end{aligned} \tag{3.5}$$

and the two expressions coincide if $x^{(1)}$ and $x^{(2)}$ are spacelike separated. The complex conjugated of the propagator, $\Delta^*(x^{(2)} - x^{(1)})$, obeys the opposite equations.

Now, by time ordering the space time points $x^{(i)}$, it can be derived, using only the Eqs. (3.5), that if amplitudes are calculated using the Feynman rules of a real Lagrange density $\mathcal{L}$ for a scalar field, and all contributing Feynman diagrams are added together, identities such as the one depicted in Fig. 1 hold. The notation here means the following: all vertices and propagators behind the shaded line are replaced by their complex conjugates (also indicated by the star in Fig. 1), and all propagators that are cut through by the shaded line are replaced by the on-shell propagator $\Delta^+(k)$, see Eq. (3.3). At the right hand side, we just have delta functions that ensure that the *in* state coincides with the *out* state. The ingoing and outgoing particles are assumed to have positive energies only. This is exactly the unitarity relation $S\,S^\dagger = \mathbb{I}$, provided that the scattering matrix amplitudes are normalized by factors $\sqrt{\varrho(m)}$. Along similar lines, dispersion relations may be derived [11].

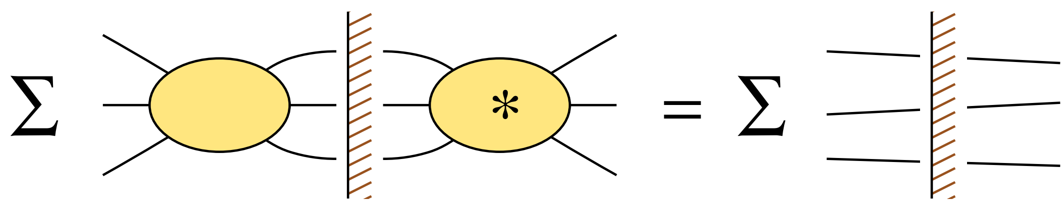


Figure 1. The unitarity relation $S\,S^\dagger = \mathbb{I}$

The functions $\varrho(m)$ for each particle species must be non negative; otherwise their square roots cannot be taken. Reversing the signs inside the square roots would lead to negative values for the ensuing probabilities, which is physically unacceptable. At this point, we can derive elegantly the conditions for a vector particle propagator to yield a unitary scattering matrix. As long as we multiply the propagators in momentum space by polynomials in k_μ, the cutting relations (3.5) continue to hold. The functions $\varrho(m)$ now include the eigenvalues of the matrix $\delta_{\mu\nu} + k_\mu k_\nu / M^2$ with of course the mass shell restriction that $k^2 + m^2 = 0$. However, we must remember that we are dealing with Minkowski space, where minus signs are associated to the time components δ_{44} of the Kronecker delta. Thus, without the $k_\mu k_\nu$ term, the matrix would have one negative eigenvalue, and unitarity would be lost. With the $k_\mu k_\nu$ term, we easily establish that there are only three positive eigenvalues, besides a vanishing one (the longitudinal particle decouples), so that unitarity is restored. Thus, according to these rules, we must keep the $k_\mu k_\nu$ term in order to ensure unitarity.

4. GHOSTS

From the previous sector one concludes that, if in the Feynman diagrams the vector fields are given propagators of the form of Eq. (2.1) then unitarity of the scattering matrix is guaranteed.

How about the propagators for *massless* vector particles, such as the photon? There are two ways to address this question. One is, to observe that the $k_\mu k_\nu$ terms in the propagators give no effect. The photon is always coupled to a conserved current $J_\mu(x)$ with $\partial_\mu J_\mu(x) = 0$. In the diagrams, one can deduce fairly easily that, as a consequence, all diagrams at the right side of the propagator cancel out when multiplied with k_μ, and all diagrams at the left cancel out when multiplied with k_ν, so, indeed, the $k_\mu k_\nu$ term multiplies two zeros. Thus, we may leave out the $k_\mu k_\nu$ term altogether.

It is slightly better, however, to observe that the photon has two polarizations, not three. [12] This should be reflected in a different numerator in the propagator. Because of gauge-invariance, this numerator is not unique. One possibility is

$$P_{\mu\nu}(k) = \frac{\delta_{\mu\nu} - (k_\mu \tilde{k}_\nu + \tilde{k}_\mu k_\nu)/2\mathbf{k}^2}{k^2}\,, \tag{4.1}$$

where $\tilde{k}_\mu$ is the vector k_μ with the sign switched in the time component, and $\mathbf{k}$ is the spacelike component of k_μ only. The numerator here only has two positive eigenvalues; the other two vanish. Unitarity here is therefore evident. Since the extra terms are proportional to k_μ or k_ν, they cancel out either because the terms at the right cancel out, or the ones at the left; this is because at both ends the propagator is coupled to a conserved current J_μ or J_ν respectively.

It was duly noted, however, that none of these arguments work for the Yang-Mills case, whether or not the Yang-Mills particles carry mass. Since these vector fields couple to themselves, one cannot say that the diagrams multiply something that is divergence free, so the $k_\mu k_\nu$ terms do not simply cancel out. Feynman [4] did discover something else: the $k_\mu k_\nu$ terms do not totally vanish, but their contributions can be replaced by an other class of diagrams, where a new kind of fictitious particle occurs. This particle is only allowed to run around in loops, but it is not allowed in the external lines. Feynman called these particles 'ghost particles'. They do not add to the spectrum of observable particles in the theory. Feynman [4] was able to prove this only for the case that his diagrams contain at most one single closed loop. Beyond that, his methods became too complicated. Veltman [6] reproduced this proof in his own way. He thought the ghost could be understood by adding to the Lagrangian a "free, non-interacting field ϕ^a", with the quantum numbers of the ghost, and unspecified mass. It was used to generate a space-time dependent (gauge-)transformation in the gauge fields A^a_μ. This did lead to identities among the Feynman diagrams, which he called 'generalized Ward identities'. But his scalar field did not behave as a ghost field as he hoped, and further manipulations were needed however to reobtain Feynman's one-loop result. These one-loop identities were beautiful, see Fig. 2, but, when he attempted to continue towards the two-loop case, he had to conclude, like Feynman, that the two-loop diagrams cannot be renormalized, since among the effective vertices extra five-point vertices and higher derivatives occur., see Fig. 3. The situation seemed hopeless.

I decided to pay full attention to Veltman's beautiful ideas, but only in as far as they were correct, such as the cutting rules and the one-loop cancellations of infinities. His introduction of "non-interacting scalar fields" to describe space-time dependent gauge transformations never made much sense to me — I suspect he thought that this is the way to give birth to ghosts in the theory — and also the mere addition of a mass term that breaks local gauge invariance could not be right. What was needed was a new analysis of the short distance structure of the theory. The Yang-Mills mass term implies that the gauge transformations will be physically observable. At short distances, the mass term would have little effect on the wild oscillations caused by these gauge transformations. This is begging for difficulties: the theory is not renormalizable.

The short distance version of the theory has effectively zero mass vector fields; therefore, I decided that a theory should be designed such that its short distance version would be of the Yang-Mills form. We *must* have an exactly gauge-invariant theory. At the same time, its long distance behavior should be such that it contains only massive vector particles. The theory that was formed in my mind is now called "Higgs theory", although, independently of P.W. Higgs [13], it

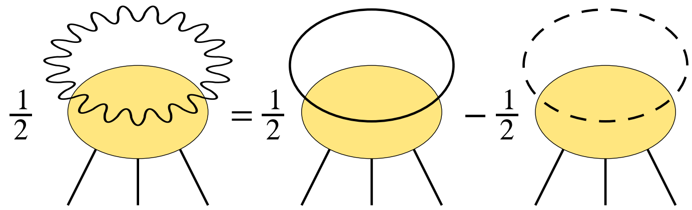
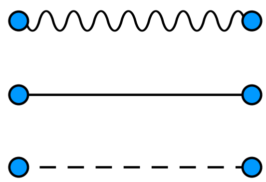


Fig. 2, At the one-loop level, the $k_\mu k_\nu$ terms in the propagator can be substituted by a renormalizable ghost contribution. Here, the shaded ovals stand for *tree* diagram insertions, and the various types of lines stand for the following propagators:



$$: \quad \frac{\delta_{\mu\nu} + k_\mu k_\nu / M^2}{k^2 + M^2 - i\varepsilon}$$

$$: \quad \frac{\delta_{\mu\nu}}{k^2 + M^2 - i\varepsilon}$$

$$: \quad \frac{1}{k^2 + M^2 - i\varepsilon}$$

was constructed and understood by F. Englert and R. Brout [14]. In a Higgs theory, gauge invariance allows us to fix the gauge, and this turns the divergent propagator (2.1) into one that is more regular at high energies.

Veltman initially was strongly opposed against theories where scalar fields develop vacuum expectation values. The reasons for his rejection of the possibility of fundamental scalar fields have always been unclear to me, until later he came with the argument of the cosmological constant. Indeed, the energy density associated with spontaneous symmetry breakdown is gigantic, and any ensuing gravitational field would cause the universe to be as curved as the surface of an orange, as he phrased it. It is my opinion, however, that questions concerning the cosmological constant must be postponed until we understand quantum gravity better.

As a compromise, it was decided that I study massless Yang-Mills theory first. This subject had been studied before by Feynman [4], B.S. DeWitt [5], L.D. Faddeev, V.N. Popov [15,16] and S. Mandelstam [17]. The Feynman rules these investigators had derived appeared to be rather abstract, and the physical significance of the theory was unclear. It did appear to be renormalizable, however. Being exactly invariant under local gauge transformations, this theory requires gauge fixing. This is achieved by the use of a Lagrange multiplier field $\lambda(x)$. However, as DeWitt, Faddeev and Popov, observed, if one uses functional integrals, a Jacobian determinant factor emerges that should not be ignored. They wrote the functional integral for the generating functions as

$$Z = \int\int \mathcal{D}A\, e^{i\int \mathrm{d}^4x\,\left(-\frac{1}{4}G^a_{\mu\nu}G^a_{\mu\nu} + \lambda^a(x)\,\partial_\mu A^a_\mu(x)\right)} \det\left(\frac{\partial A^a_\mu(x)}{\partial\Lambda^b(x')}\right) \,. \tag{4.2}$$

A beautiful short paper by Faddeev and Popov [15,16] explained this result, but it did not mention how to calculate the determinant. This is in their more extended article, which unfortunately arrived in the Western World only much later. But it was not hard to reconstruct [12]. One carries out the functional integral

$$\det{}^{-1}\left(\mathcal{M}^a_b(x,x')\right) = C\int\int \mathcal{D}\eta\mathcal{D}\overline{\eta}\, e^{i\int \mathrm{d}^4x\,\mathrm{d}^4x'\,\left(\overline{\eta}_a(x)\mathcal{M}^a_b(x,x')\eta^b(x')\right)} \,. \tag{4.3}$$

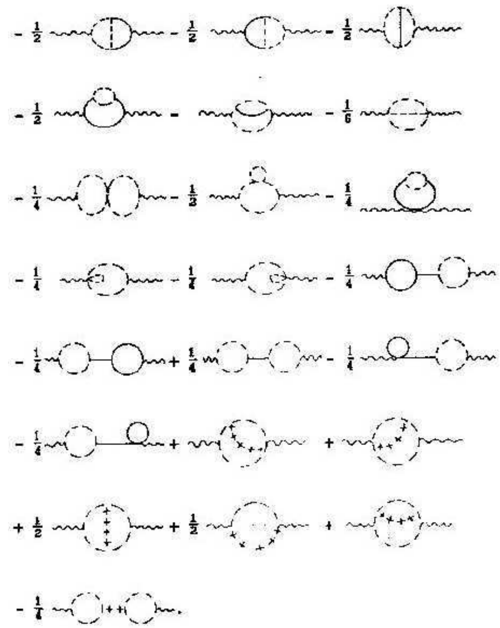


Fig. 3. Extra vertices obtained by Veltman at the two-loop level. From Ref [6]. These would be the diagrams of the two-loop contributions to the propagator insertions. Although the ghost contributions make the associated expressions less divergent than in the manifestly unitary representation, the theory is still non-renormalizable because of extra momentum vertices in the extra vertices, some of which also connect too many lines.

Since the logarithm of this expression generates the one-loop diagrams, it is easy to see that the inverse sign (-1) can be cancelled by attributing an extra minus sign to each closed loop of 'η particles'. This turns the η particles into fermions, in spite of the fact that they are scalars. One can also note that treating the η fields as anticommuting Grassmann fields directly produces the determinant as in Eq. (4.2) and not its inverse.

The Lagrangian (4.2) does not yet exactly produce the simple propagator $\delta_{\mu\nu}/(k^2 - i\varepsilon)$ for the Yang-Mills bosons; it yields a term of the form $k_\mu k_\mu/k^4$. A modest trick in the functional integral removes this unwanted term.

The η field corresponds to a ghost particle, but there was an important difference with the Feynman–Veltman ghost: a factor 2. This is because the η field is complex, and the propagator is henceforth an oriented one. Thus, relating this result to the problem of massive vector particles would be a non-trivial one. How can one see that, using these rules, the diagrams would be unitary?

5. WARD IDENTITIES

The technique to use a space-time dependent transformation, or 'a transformation generated by the field of a free particle' was leading me nowhere. It only generated propagator identities of the form of Fig. 4. This was far from sufficient to derive unitarity. The functional integrals provided *formal* arguments for the theory to be unitary, but how could one check this at the level of diagrams? The question was important in particular because we needed to know how to replace divergent integrals by finite ones; this we would only be able to do explicitly for Feynman diagrams. How does one replace Veltman's expression of Fig. 2? Something like this had to be valid for diagrams with arbitrarily many loops in them.

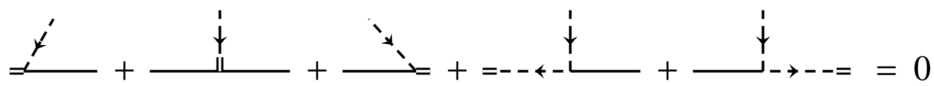


Fig. 4. identity relating vertices and propagators for the massless Yang-Mills theory.

In fact, what was needed was a deductive procedure to reiterate identities such as Fig. 4 inside a diagram. But then, sooner or later, the ghost lines would meet themselves, and what then? A new identity was needed, see Fig. 5. It is *not* related to a symmetry, but to the Jacobi identity of the gauge group:

$$f_{aib}f_{bjk} + f_{ajb}f_{bki} + f_{akb}f_{bij} = 0\,. \tag{5.1}$$

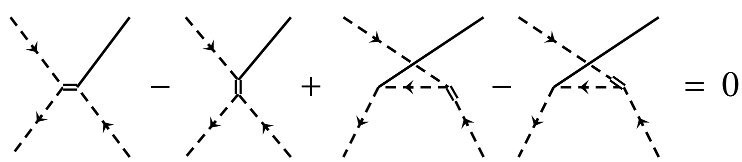


Fig. 5. identity for ghosts meeting ghosts. The minus signs here could not be deduced from a symmetry argument (or so it seemed, but see Sect. 7.)

Now, to prove unitarity, I did exactly as was asked by Veltman: prove Ward identities for the theory on the mass shell. If all external lines of a diagram are on the mass shell, but some of them longitudinally polarized, one gets the identities of Fig. 6. This enabled us to show the exact cancellation, at all orders of the longitudinally polarized particles against the ghosts in the intermediate states, so that unitarity follows: [12]

$$\sum_n S|n\rangle\langle n|S^\dagger = \mathbb{I}\,, \tag{5.2}$$

where the intermediate states $|n\rangle$ only include transverse vector particles.

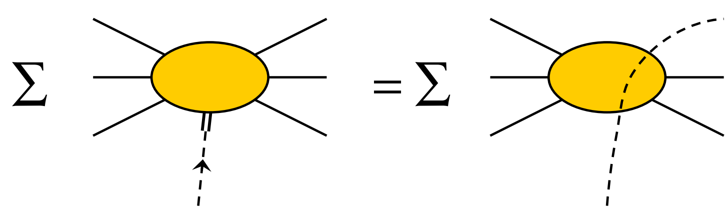


Fig. 6. Ward identity for on shell diagrams.

Shortly after this, A.A. Slavnov [18] and J.C. Taylor [19] independently pointed out that the Ward identities can be generalized to off the mass shell. It filled me with pride that they both referred to my paper and they used my result, although for them the much simpler functional integral argument sufficed. I did warn later investigators that the off-shell identities would involve new kinds of divergences that would require independent renormalizations — a fact that one can more fully appreciate when the combinatorial arguments in the Feynman diagrams are applied. Perhaps I should also have objected when Stora's group began to use what they called the 'Slavnov-Taylor identities' [20]; my name could have been added, but at that time I was not in the least concerned or even interested in priority issues.

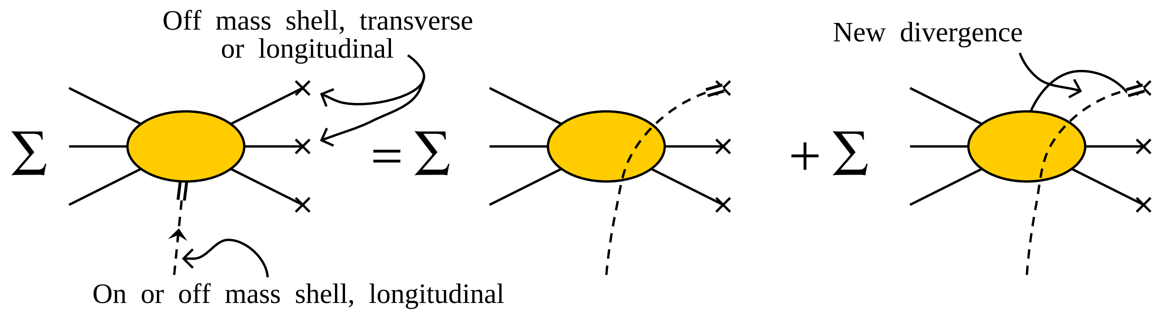
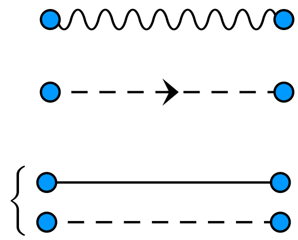


Fig. 7. Slavnov-Taylor identity for off shell diagrams.

6. THE MASSIVE CASE

In the mean time I had the fortune to attend an excellent summer school. At Cargèse, Corsica, at a beach resort, M. Lévy had established an Institute of Science, and in the summer of 1970, the renormalizability of the so-called sigma model [21] for the strong interactions would be discussed. Now this was a quite interesting model for strongly interacting mesons and baryons. It exhibited a global chiral symmetry, *which was spontaneously broken.* One topic of investigation consisted of various attempts to tame the highly divergent nature of the perturbation series of such models, in which the perturbation expansion parameter, the strong coupling constant g, would *not* be small. Resummation techniques such as the Padé expansion were thought to be applicable here, but, as it is known now, they would fail. The careful analysis of the renormalization procedure itself, by B.W. Lee [22], K. Symanzik [23], J.-L. Gervais [22] and others, was much more interesting. They established that the renormalization counter terms would not be affected by spontaneous symmetry breaking; they were in fact the same as the counter terms of the symmetric theory, and the underlying symmetry properties of the theory would not be affected.

Now, only global gauge symmetries were discussed at Cargèse. To me, the step from global to local symmetries did not seem to be such a drastic one, but when I asked Lee and Symanzik about this, they referred to Veltman. Since Veltman was still rejecting theories with vacuum expectation values, I realized that I was on my own here, but what I had heard at Cargèse strongly encouraged me to continue investigating the ideas that I had about the role of scalar fields in a Yang-Mills theory. If you write a scalar field ϕ as $\phi = F + \tilde{\phi}$, then $\tilde{\phi}$ may be considered as a perturbative correction. If the perturbation expansion is carried out from this viewpoint, the propagators of the theory will correspond to the following expressions,



$$
\begin{aligned}
&\text{gauge photon,} \quad : \quad \frac{\delta_{\mu\nu}}{k^2+M^2-i\varepsilon}\ ; \\
&\text{FP ghost,} \quad : \quad \frac{1}{k^2+M^2-i\varepsilon}\ ; \\
&\text{Higgs,} \quad \left\{ \begin{array}{l} \\ \\ \end{array} \right. : \quad \begin{array}{c} \dfrac{1}{k^2+M_H^2-i\varepsilon} \\ \dfrac{1}{k^2+M^2-i\varepsilon} \end{array} .
\end{aligned}
\qquad (6.1)
$$

The mass term here arises from the term $(D_\mu F)^2$, and, in combination with the other interactions it is now completely gauge-invariant.

Notice that the Higgs field generates a ghost propagator without an orientation in it (since the field is a real one, not complex). This ghost field simply consists of those components of the Higgs field that can be gauge-transformed away. Consequently, a closed loop of the Higgs ghost cancels exactly one half of the Faddeev-Popov closed loop (which carried its fermionic minus sign, but a weight 2 because of its being oriented). This would explain why there had been a factor 2 discrepancy between the ghost prescription of Feynman and Veltman for the massive case on the one hand, and Faddeev, Popov, DeWitt and Mandelstam's prescription on the other. The latter had restricted themselves to the strictly massless Yang-Mills theory.*

I had to complete my work on the massless case first. Unitarity, finiteness and uniqueness were discussed at length with Veltman, discussions which were instrumental for the production of my first paper [12], but he still thought that the massive case would be an entirely different piece of cake. Eventually I could convince him by presenting the Lagrangian in the shifted notation, in terms of the fields $\tilde{\phi}$, where the vacuum expectation values are suppressed. My paper [24] on the massive Yang-Mills theory† was presented to *Nuclear Physics B* in June 1971, just at the time when an International Conference at Amsterdam gave me the opportunity for a (very brief) presentation.

With the Cargèse school in mind, where feverish attempts had been made at understanding the hadron spectrum, and where incorporating the ϱ vector meson had been regarded as a distant target, I thought that massive Yang-Mills theories should be applied there, but Veltman, and others, quickly convinced me of their importance for understanding the weak force. I had already reproduced Weinberg's lepton model [25] before I became aware of his paper.

---

* The Feynman rules for the Faddeev-Popov ghost and those for the unphysical Higgs component are not the same, yet they do have the same effects at the one-loop level. This is why one can view the $-\frac{1}{2}$ in Fig. 2 as $-1+\frac{1}{2}$.

† In this paper, a property of the massless theory was alluded to that would later be called 'asymptotic freedom'.

7. EPILOGUE.

After these developments, much work remained to be done. The years that followed were Golden Years for Elementary Particle Theory, certainly from my perspective. First, our regulator technique, which until then only guaranteed anomaly free regularization at the one-loop level, had to be improved. Could my five-dimensional approach be generalized? Unfortunately not. Eventually, a method was found: we should use $4-\varepsilon$ dimensions, not an integral number of them. This was an audacious step. I still have a first draft of a manuscript proposing this idea, and of which I am the sole author, but together with Veltman, the method was further refined, and this led to our joint publication of this proposal to use fractional dimensions to regularize gauge theories to all orders in the coupling constant. [26]

Dimensional regularization failed when chiral fermions come into play, since the chiral projection can only be carried out in four space-time dimensions. Exactly here, however, the occurrence of the Adler-Bell-Jackiw anomaly [9] causes incurable diseases: no regulator method of whatever form can possibly work here; one has to cancel the anomalies out. [27] If the anomalies do cancel at the one-loop level, it can be shown that no further clashes will occur, even non-perturbatively. It is my opinion that this can be seen rather easily, [28] and the result is not disputed, but others do claim that a complete proof of this statement is quite difficult and delicate. [29]

A drastic improvement of the renormalizability proof was discovered by Becchi, Rouet and Stora: the Slavnov-Taylor identities can be derived from a symmetry after all, but the symmetry is an anticommuting one, a supersymmetry: the fermionic Faddeev-Popov ghost fields, η and $\overline{\eta}$, can be transformed into the bosonic, but unphysical, gauge dependent field components of the theory. The generator of this *global* supersymmetry is an anticommuting constant $\overline{\varepsilon}$. Let, in the simplest case, the Lagrangian be written in the shorthand notation:

$$\mathcal{L}(A,\eta,\overline{\eta}) = \mathcal{L}^{\rm inv}(A) + \lambda^a(x)C^a(A,x) + \overline{\eta}^a(x)\frac{\partial C^a(x)}{\partial\Lambda^b(x')}\eta^b(x')\,, \tag{7.1}$$

Here, the function $C^a(x)$ is the field combination used to fix the gauge, for instance $\partial_\mu A_\mu(x)$, and $\lambda^a(x)$ is a Lagrange multiplier field. Then the BRST symmetry operation is written as

$$\begin{aligned}
\delta A^a(x) &= \overline{\varepsilon}\frac{\partial A^a(x)}{\partial\Lambda^b(x')}\eta^b(x')\,;\\
\delta\eta^a(x) &= \tfrac{1}{2}\overline{\varepsilon}f^{abc}\eta^b(x)\eta^c(x)\,;\\
\delta\overline{\eta}^a(x) &= -\overline{\varepsilon}\lambda^a(x)\;;\\
\delta\lambda^a(x) &= 0\,.
\end{aligned} \tag{7.2}$$

It is not difficult to generalize this to the case where we choose the gauge to be fixed by a quadratic function of $C^a(x)$, such as $(C^a)^2$. The action S is invariant under this symmetry. An essential requirement for the invariance proof is again the Jacobi identity (5.1).

This result dispenses of the need for lengthy combinatorial proofs of the Slavnov-Taylor identities. Regularization is henceforth required to respect BRST symmetry. Unitarity and other desired features can now be shown by first demonstrating that the physical (that is, gauge-invariant) components of the scattering amplitudes do not depend on the choice of the function $C^a(x)$ due to this symmetry.

With the tools of BRST quantization an dimensional renormalization, the perturbative treatment of renormalization has now become a matter of routine. Since then, many successful attempts were made towards better understanding of gauge theories beyond the perturbation expansion, such as the large N approximation, the renormalization group, instantons, Monte Carlo simulations on a lattice, analytic approaches to supersymmetric versions of the theory, and more.

REFERENCES

1. See the arguments raised in: R.P. Feynman, in *The Quantum Theory of Fields – The 12th Solvay Conference*, 1961 (Interscience, New York).
2. T.Y. Cao and S.S. Schweber, *The Conceptual Foundations and Philosophical Aspects of Renormalization Theory*, *Synthese* **97** (1993) 33, Kluwer Academic Publishers, The Netherlands.
3. C.N. Yang and R.L. Mills, *Phys. Rev.* **96** (1954) 191, *see also:* R. Shaw, Cambridge Ph.D. Thesis, unpublished.
4. R.P. Feynman, *Acta Phys. Polonica* **24** (1963) 697.
5. B.S. DeWitt, *Phys. Rev.* Lett. **12** (1964) 742, and *Phys. Rev.* **160** (1967) 1113; *ibid.* **162** (1967) 1195, 1239.
6. M. Veltman, *Physica* **29** (1963) 186, *Nucl. Phys.* **B7** (1968) 637; J. Reiff and M. Veltman, *Nucl. Phys.* **B13** (1969) 545; M. Veltman, *Nucl. Phys.* **B21** (1970) 288.
7. R.P. Feynman and M. Gell-Mann, *Phys. Rev.* **109** (1958) 193; see the earlier work by E.C.G. Sudarshan and R.E. Marshak, Proc. Padua-Venice Conf. on *Mesons and Recently discovered Particles*, 1957, p. V - 14, reprinted in: P.K. Kabir, *Development of Weak Interaction Theory*, Gordon and Breach, 1963, p. 118; E.C.G. Sudarshan and R.E. Marshak, *Phys. Rev.* **109** (1958) 1860.
8. S.L. Glashow, *Nucl. Phys.* **22** (1961) 579.
9. S.L. Adler, *Phys. Rev.* **177** (1969) 2426; J.S. Bell and R. Jackiw, *Nuovo Cim.* **60A** (1969) 47. Also see: H. Fukuda and Y. Miyamoto, Progr. Theor. Phys. **4** (1949) 347.
10. N. Bogolyubov and D. Shirkow, *Fortschr. der Physik* **3** (1955) 439; N. Bogolyubov, B. Medvedev and M. Polivanov, *Fortschr. der Physik* **5** (1958) 169.
11. G. 't Hooft and M. Veltman, "DIAGRAMMAR", CERN Report 73/9 (1973), reprinted in "Particle Interactions at Very High Energies, NATO Adv. Study Inst. Series, Sect. B, vol. 4b, p. 177.
12. G. 't Hooft, *Nucl. Phys.* **B33** (1971) 173.
13. P.W. Higgs, *Phys. Lett.* **12** (1964) 132; *Phys. Rev.* Lett. **13** (1964) 508; *Phys. Rev.* **145** (1966) 1156.
14. F. Englert and R. Brout, *Phys. Rev. Lett.* **13** (1964) 321.
15. L.D. Faddeev and V.N. Popov, *Phys. Lett.* **25B** (1967) 29.
16. L.D. Faddeev, *Theor. and Math. Phys.* **1** (1969) 3 (in Russian), *Theor. and Math. Phys.* **1** (1969) 1 (Engl. transl).

17. S. Mandelstam, *Phys. Rev.* **175** (1968) 1580, 1604.

18. A.A. Slavnov, *Theor. Math. Phys.* **10** (1972) 153 (in Russian), *Theor. Math. Phys.* **10** (1972) 99 (Engl. Transl.) 31.

19. J.C. Taylor, *Nucl. Phys.* **B33** (1971) 436.

20. C. Becchi, A. Rouet and R. Stora, *Commun. Math. Phys.* **42** (1975) 127; id., *Ann. Phys.* (N.Y.) **98** (1976) 287; see also I.V. Tyutin, Lebedev Prepr. FIAN 39 (1975), unpubl.

21. M. Gell-Mann and M. Lévy, *Nuovo Cim.* **16** (1960) 705.

22. B.W. Lee, *Nucl. Phys.* **B9** (1969) 649; J.-L. Gervais and B.W. Lee, Nucl. Phys. **B12** (1969) 627; B.W. Lee, "Chiral Dynamics", Gordon and Breach, New York, 1972, ISBN 0 677 01380 9 (cloth); 0 677 01385 X (paper).

23. K. Symanzik, Let. Nuovo Cim. **2** (1969) 10, *Commun. Math. Phys.* **16** (1970) 48.

24. G. 't Hooft, *Nucl. Phys.* **B35** (1971) 167.

25. S. Weinberg, *Phys. Rev. Lett.* **19** (1967) 1264.

26. G. 't Hooft and M. Veltman, *Nucl. Phys.* **B44** (1972) 189; *Nucl. Phys.* **B50** (1972) 318. See also: C.G. Bollini and J.J. Giambiagi, *Phys. Letters* **40B** (1972) 566; J.F. Ashmore, *Nouvo Cim. Letters* **4** (1972) 289.

27. C. Bouchiat, J. Iliopoulos and Ph. Meyer, Phys. Let. **38B** (1972) 519.

28. G. 't Hooft, *Physics Lett.* **B349** (1995 491.

29. M. Luescher, in Erice lecture notes, 38th Course, *Theory and Experiment Heading for new Physics*, Aug.—Sept. 2000.