FROM WEAK INTERACTIONS TO GRAVITATION

by

Martinus J. G. Veltman

Emeritus MacArthur Professor of Physics, University of Michigan, Ann Arbor, MI 48109-1120, USA.

INTRODUCTION

This lecture is about my contribution to the proof of renormalizability of gauge theories. There is of course no perfectly clear separation between my contributions and those of my co-laureate 't Hooft, but I will limit myself to some brief comments on those publications that carry only his name. An extensive review on the subject including more detailed references to contemporary work can be found elsewhere [1].

As is well known, the work on the renormalizability of gauge theories caused a complete change of the landscape of particle physics. The work brought certain models to the foreground; neutral currents as required by those models were established and the discovery of the J/Ψ was quickly interpreted as the discovery of charm, part of those models as well. More precisely, we refer here to the model of Glashow [2], the extension to include quarks by Glashow, Iliopoulos and Maiani (GIM) [3] and the model of Weinberg-Salam [4] for leptons including a Higgs sector. The GIM paper contained discussions on the required neutral hadron currents and also the inclusion of charm as suggested first by Hara [5]. After an analysis by Bardeen, at a seminar in Orsay (see also [6]), the work of Bouchiat, Iliopoulos and Meyer [7] established the vanishing of anomalies for three color quarks. Without going into details, subsequently quantum chromodynamics came to be accepted. In this way the Standard Model was established in just a few years.

ESSENTIAL STEPS

Let me review here what I consider as my contribution to the subject. It can be described in three separate parts. I will try to simplify things as much as possible.

I. **The physics argument.** In 1965 Adler [8] and Weisberger [9] established what is now known as the Adler-Weisberger relation. This relation, numerically agreeing with experimental data, was interpreted by me as a consequence of a Ward identity of a non-Abelian gauge theory (also called Yang-Mills theories), and as such guided me to take up the study of such theories.

II. **The renormalizability argument.** Earlier calculations on the radiative corrections to the photon-vector-boson vertex showed a disappearance of many divergencies for a properly chosen vector-boson magnetic moment. In studying Yang-Mills theories I noted that those theories automatically produced this particular magnetic moment. I therefore concluded that Yang-Mills theories are probably the best one can have with respect to renormalizability. Thus I was led to the study of renormalizability of these theories.

III. **Technical progress.** Starting then on the study of diagrams in a Yang-Mills theory I established the vanishing of many divergencies provided the external legs of the diagrams were on the mass shell. That by itself is not enough with respect to renormalizability, because that requires diagrams and Feynman rules of a renormalizable type. I was thus led to search for a transformation of the theory such that new, renormalizable type Feynman rules were derived, without changing the S matrix. In this I succeeded up to one loop.

None of these points is trivial, as can be shown easily by considering work in that period. For example, Weinberg [10] in his 1979 Nobel lecture reports that he interpreted the success of the Adler-Weisberger relation as a property of strong interactions, namely, the validity of chiral SU2 × SU2. Consequently he continued working on things such as π–π scattering. Feynman is reputed to have exclaimed that he never thought of investigating the renormalizability aspect of Yang-Mills theories when he heard of that development. Finally, there existed several papers where the non-renormalizability of Yang-Mills theories was "proven", for example one by Salam [11].

In the following I will discuss these three points in detail as they developed historically.

THE PHYSICS ARGUMENT

In 1965 Adler and Weisberger derived their famous relation between the axial vector coupling constant of β decay in terms of a dispersion integral for pion-nucleon scattering. This relation, agreeing well with experiment, was based on Gell-Mann's current commutator rules [12]. Subsequently an extensive discussion developed in the literature concerning the so-called Schwinger terms that could invalidate the argument. I decided to try to derive these same results starting from another assumption, and as a starting point I took the well-known CVC and PCAC equations for the weak currents:

$$\partial_\mu J_\mu^V = 0$$

$$\partial_\mu J_\mu^A = ia\pi$$

These equations do not include higher-order electromagnetic (e.m.) or weak effects. As a first step I tried to include electromagnetic effects by using the well-known substitution $\partial_\mu \to \partial_\mu - iqA_\mu$, where q is the charge of the object on which ∂_μ operates. Since the currents were isovectors, that could be done easily using isospin notation. For the vector current the equation became

$$(\partial_\mu + ie\vec{A}_\mu \times)\vec{J}^V_\mu = 0,$$

treating the e.m. field as the third component of an isovector. Next I used the idea that the photon and the charged vector bosons may be seen as an iso-triplet, thus generating what I called divergency conditions [13]. For the axial vector current this gave:

$$\partial_\mu \vec{J}^A_\mu = ia\vec{\pi} + ie\vec{A}_\mu \times \vec{J}^A_\mu + ig\vec{W}^V_\mu \times \vec{J}^A_\mu + ig\vec{W}^A_\mu \times \vec{J}^V_\mu.$$

As a matter of technical expedience two vector bosons were used to denote a vector-boson coupling to a vector current and a vector-boson coupling to an axial current. This equation turned out to be adequate to derive the Adler-Weisberger relation. An added benefit of this derivation was that no difficulties with respect to Schwinger terms arose, and the axial vector coupling constant was related directly to the pion-nucleon scattering length. The Adler-Weisberger relation evidently used an additional relation in which the pion-nucleon scattering length was given in terms of a dispersion integral.

In response, John Bell, then at CERN, became very interested in this derivation. He investigated what kind of field theory would generate such divergency conditions, and he found that this would happen in a gauge theory [14].

Subsequent developments were mainly about the consequences of those relations involving e.m. fields only. It is clear that specializing to the third component of the axial divergence condition there are no e.m. corrections. Following Adler, reading $\partial_\mu J^{A0}_\mu = a\pi^0$ in the opposite way imposed a condition on the pion field including e.m. effects. This was an extension of earlier work by Adler [15], known under the name of consistency conditions for processes involving pions. In this case one of the conclusions was that π^0 decay into two photons was forbidden, and without going into a detailed description this led to the work of Bell and Jackiw [16] on the anomaly. Simultaneously Adler [17] discovered the anomaly and in fact used precisely my (unpublished) derivation to connect this to π^0 decay. Later I became quite worried by this development, as I saw this anomaly as a diffculty with respect to renormalization.

THE RENORMALIZABILITY ARGUMENT

Here I must go back to 1962. In that year Lee and Yang [18] and later Lee [19] alone started a systematic investigation of vector bosons interacting with photons. The paper of Lee and Yang mainly concentrated on deriving the Feynman rules for vector bosons. The trouble at that time was that in doing the usual canonical derivation one encountered certain contact terms for the vector-boson propagator. I will not delve any further into this; later I found a simple way to circumvent these problems. In those days, however, these were considered serious problems.

Subsequently Lee started a complicated calculation, namely the lowest or-

der radiative corrections to the vector-boson-photon coupling. The usual replacement $\partial_\mu \rightarrow \partial_\mu - ieA_\mu$ in the vector-boson Lagrangian is not suffcient to determine the vector-boson magnetic moment; it remains an arbitrary parameter. This is because of the occurrence of two derivatives such as $\partial_\mu\partial_\nu$; when making the minimal substitution it matters whether one writes $\partial_\mu\partial_\nu$ or $\partial_\nu\partial_\mu$, and this causes the arbitrariness in the magnetic moment. Anyway, Lee, concentrating on the electric quadrupole moment of the vector-boson calculated the appropriate triangle diagram using a cutoff procedure called the ξ-limiting process.

I was very interested in this calculation, because like many physicists I strongly believed in the existence of vector bosons as mediators in weak interactions. This belief was based on the success of the $V - A$ theory, suggesting a vector structure for the weak currents. Indeed, this led Glashow to his famous 1961 paper. I decided that Lee's work ought to be extended to other situations, but it was quite obvious that this was no mean task. Given the ξ method, and the occurrence of the magnetic moment as an arbitrary parameter, the triangle diagram if calculated fully (Lee limited himself to parts relevant to him) gives rise to a monstrous expression involving of the order of 50 000 terms in intermediate stages. There was simply no question of going beyond the triangle diagram.

At this point I decided to develop a computer program that could do this work. More specifically, I concentrated on the triangle graph, but I wrote the program in such a way that other processes could be investigated. In other words, I developed a general purpose symbolic manipulation program. Working furiously, I completed the first version of this program in about three months. I called the program *Schoonschip,* among others to annoy everybody not Dutch. The name means "clean ship," and it is a Dutch naval expression referring to clearing up completely a messy situation. In January 1964, visiting New York in connection with an American Physical Society meeting, I visited Lee and told him about the program. He barely reacted, but I heard later that after I left the office he immediately wanted one of the local physicists to develop an analogous program.

, In toying with the calculation I tried to establish what would be the best value for the vector-boson magnetic moment with respect to the occurring divergencies. There was one value where almost all divergencies disappeared, but I did not know what to do with this result. It remained in my memory though, and it played a role as explained below.

TECHNICAL PROGRESS

To explain the development requires some backtracking. In 1959 I took up the study of the problem of unstable particles. The problem is of a nonperturbative character, because a particle is unstable no matter how small the coupling constant of the interaction that produces the decay. Thus the (unstable) particle will not appear in the in and out states of the S matrix. However, for zero value of the coupling constant the particle is stable and must be

part of the in and out states. Thus the limit of zero coupling constant does not reproduce the zero-coupling-constant theory.

It was in principle well known at the time how to handle an unstable particle. Basically one did what is called a Dyson summation of the propagator, and that indeed removed the pole in the propagator. From the Källén-Lehmann representation of the propagator one knows that every pole corresponds to an in or out state, so the summation indeed seemed to correspond to removing the particle from the in and out states.

However, when performing the Dyson summation one found that the theory became explicitly nonperturbative, as self-energy diagrams and with them factors g (the coupling constant of the destabilizing interaction) appeared in the denominator of the propagator of the unstable particle. That propagator looked like this:

$$\frac{1}{k^2 + M^2 + g^2 F(k).}$$

Obviously, this propagator cannot be expanded as a function of g in the neighbourhood of $k^2 + M^2 = 0$ if the imaginary part of $F(k)$ is nonzero in that point (the real part is made zero by mass renormalization). So, instead of a propagator with a pole the Dyson summation made it into a function with a cut in the complex k^2 plane. At this point it is no longer clear that the S matrix is unitary, because the usual equation for the S matrix $S = T[exp(iH)]$ is no longer valid. In other words, to establish unitarity one had to consider the diagrams by themselves.

Thus I attacked this problem, essentially finishing it in 1961. This was for my thesis, under the supervision of Leon van Hove. The article went its ponderous Dutch thesis way and was published in 1963 [20], in a somewhat unusual journal (*Physica*) for high-energy physics. Curiously, about the same time Feynman [21] considered the same problem, in connection with establishing unitarity for the massless Yang-Mills theory, a theory whose diagrams include ghosts. These ghosts make unitarity nonevident. Moreover the derivation by Feynman, done with path integrals, did not guarantee unitarity. I am quite sure that he never saw my article, and I never discussed it with him either. He tried to do it some other way, quite complicated, initially succeeding only up to one loop. Later, DeWitt [22] extended Feynman's proof to any number of loops, but my proof is much simpler and moreover connects quite directly to physical intuition. In fact, my proof had as a result that the imaginary part of a diagram equals the sum of all diagrams that can be obtained by cutting the initial diagram in all possible ways.

The importance of this work was twofold. Not only did unitarity become a transparent issue, I also learned to consider diagrams disregarding the way they were derived, for example using the canonical formalism. Given that it is not easy to derive Feynman rules for a Yang-Mills theory in the canonical way, that did gave me an advantage in studying that theory. For considering Yang-Mills theories the path-integral formalism is quite adequate; there is only one point, and that is that this formalism does not guarantee unitarity. In

1968 the path-integral formalism had all but disappeared from the literature, although students of Schwinger did still learn functional methods. I myself did not know the first thing about it.

In 1968, I was invited by Pais to spend a month at Rockefeller University. I happily accepted this invitation and decided to try to think through the present situation. For two weeks I did nothing but contemplate the whole of weak interactions as known at the time. I finally decided to take Bell's conclusion seriously and therefore assumed that the weak currents were those of a gauge theory. Thus I started to learn Yang-Mills theory and tried to find out how that would work in some simple weak processes. In the process of writing down the Feynman rules I noted that this theory gave precisely the "best" vertex (with respect to divergencies) as I had found out doing the work reported above on the photon-vector-boson interaction. This encouraged me to concentrate on the renormalizability aspect of the theory.

As far as I remember I started by considering the one-loop corrections to neutrino-electron scattering. Here the situation became quickly quite complicated. The vertices of the Yang-Mills theory were much more complicated than those that one was used to, and even the simplest diagrams gave rise to very involved expressions. In the end I decided to drop everything except the basic theory of vector bosons interacting with each other according to a Yang-Mills scheme. In addition, of course, I gave these vector bosons a mass, since the vector bosons of weak interactions were obviously massive. I started in blissful ignorance of whatever was published on the subject, which was just as well or I might have been convinced that Yang-Mills theories are non-renormalizable. As Feynman said in his Nobel lecture as presented at CERN, "Since nobody had solved the problem it was obviously not worthwhile to investigate whatever they had done." I want to mention here that at that time I started to get worried about the anomaly, but I decided to leave that problem aside for the moment.

Consider the propagator for a massive vector field:

$$\frac{\delta_{\mu\nu} + k_\mu k_\nu / M^2}{k^2 + M^2} .$$

The source of all trouble is of course the $k_\mu k_\nu$ term. So anyone starting at this problem tries to eliminate this term. In quantum electrodynamics that can indeed be done, but for Yang-Mills fields this is not possible. There is always some remnant. Now here a simple observation can be made: this bad term comes with a factor $1/M^2$. In fact, one can trace the worst divergencies in a diagram simply by counting factors $1/M^2$. But given that they will not ever completely cancel, as they indeed do not for a Yang-Mills-type theory, then one will never get rid of these divergencies unless somehow these factors $1/M^2$ cancel out. But where should the necessary factors M^2 come from? There is only one way, and that is through external momenta that are on the mass shell, meaning that the momentum p of such an external line satisfies $p^2 = -M^2$. And here is the problem.

Renormalization means that for a divergent graph one cannot take the ex-

ternal momentum on the mass shell and then do the necessary subtractions, because the graph may occur as part of a more complicated graph. For example, in a box diagram there may be a self-energy insertion in one of the internal lines. There momenta of the lines attached to the self-energy insertion are not on the mass shell, thus it is not suffcient to subtract only those divergencies that remain if those momenta are on the mass shell. You would still have to show that the extra divergencies arising when those lines are not on the mass shell actually cancel, a gruesome task. What to do?

Well, what I did was to reformulate the theory such that somehow all cancellations were implemented in the rules. In the first instance I took the Stueckelberg technique [23]: added a scalar field and made couplings involving derivatives such that it appeared together with the vector-boson propagator:

$$\frac{\delta_{\mu\nu} + k_\mu k_\nu/M^2}{k^2 + M^2} + \kappa \; \frac{k_\mu k_\nu/M^2}{k^2 + M^2}.$$

The second term would be due to the exchange of the scalar particle. The parameter κ was introduced to keep track of the counter term. Eventually κ was taken to be -1. Now this new field is physically undesirable, because to have $\kappa = -1$ is actually impossible. The scalar field would have to have indefinite metric, or some such horrible thing. In an Abelian theory it is easy to show that the field is noninteracting, but not in the non-Abelian case. Then I had an idea: introduce further interactions of this new scalar field in such a way that it becomes a free field. The result, hopefully, would be a new theory, involving a well-behaved vector-boson propagator and furthermore an interacting scalar field that would then be a ghost. Indeed, being a free field it could appear in the final state only if it was there in the initial state. At this point one would have new Feynman rules, presumably much less divergent because of the improved vector-boson propagator. It was all a matter of what Feynman rules would result for this scalar field. If they were those for a renormalizable field, then we would be in business!

So here is the important point: the theory must be formulated in terms of diagrams which would have to be of the renormalizable type. No matter that ghosts occur; those do not get in the way with respect to the renormalization program.

There is a bonus to this procedure: one can write an amplitude involving one such scalar field. Because the scalar field is a free field, that amplitude must be zero if all other external lines are on the mass shell. That then gives an identity. Using Schwinger's source technique one can extend this to the case in which one or more of the other external lines are off the mass shell. I later called the resulting identities generalized Ward identities.

There is another aspect to this procedure. Because the final diagrams contain a vector-boson propagator that has no $k_\mu k_\nu$ part, that theory is not evidently unitary. At that point one would have to use the cutting rules that I had obtained before and show, using Ward identities for the cut scalar lines, that the theory was unitary. All in all quite a complicated affair, but not that diffcult.

Here the miracle occurred. On the one-loop level almost all divergencies disappeared. It was not as straightforward as I write it here, because even with the new rules one needed to do some more work using Ward identities to get to the desired result. In any case, I arrived at Feynman rules for one-loop diagrams that were by ordinary power-counting rules renormalizable rules. For those who want to understand this in terms of the modern theory: instead of a Faddeev-Popov ghost (with a minus sign for every closed loop) and a Higgs ghost (no minus sign, but a symmetry factor) I had only one ghost, and on the one-loop level that was actually the difference of the two ghosts as we know them now.

No one will know the elation I felt when obtaining this result. I could not yet get things straight for two or more loops, but I was sure that that would work out all right. The result was for me a straight and simple proof that my ideas were correct. A paper presenting these results was published [24].

The methods in that paper were clumsy and far from transparent or elegant. The ideas, however, were clear. I cannot resist quoting the response of Glashow and Iliopoulos [25]. After my paper appeared they decided to work on that problem as well, and indeed, they showed that many divergencies cancelled, although not anywhere to the renormalizable level. For example, the one-loop box diagram is divergent like Λ^8 in the unitary gauge; their paper quoted a result of Λ^4. I of course obtained the degree of divergence of a renormalizable theory, i.e., $\log(\Lambda)$. Here then is a part of the footnote they devote to this point: "The divergencies found by M. Veltman go beyond the theorem proven in this paper, but they only apply to on-mass-shell amplitudes..." Indeed!

In present-day language one could say that I made a transformation from the unitary gauge to a "renormalizable" gauge. As I had no Higgs the result was not perfect. But the idea is there: there may be different sets of Feynman rules giving the same S matrix.

INTERIM

In the years 1969–1971 I expended considerable effort trying to go beyond the one-loop result. There were many open problems, and they had to be considered. In the beginning of 1970 I streamlined the derivation to the point that it became transparent. This was done by deriving Ward identities using Schwinger's source techniques [26]. This is really much like the way one derives Ward identities today, now called Slavnov-Taylor identities. The BRS transformation is a sophisticated form of the free-field technique (using anticommuting fields). I remember being upset when I first heard a lecture by Stora on what he called the Slavnov-Taylor identities. I told him that they were another variant of my generalized Ward identities. However, Stora is not a diagram man, and I am sure that he never understood my paper.

Another issue was the limit of zero mass of the massive Yang-Mills theory. In January 1969 there was a conference at CERN and I announced that two-loop diagrams for the massive Yang-Mills theory did not go over into those of the

massless theory, in other words, the massless theory is not the limit of zero mass of the massive theory [27]. This argument was spelled out in an article with J. Reiff [28]. The main part of that article was to tie up another loose end: the Feynman rules for vector bosons in the unitary gauge. The argument is quite elegant and superseded the article by Lee and Yang mentioned before. This made it clear to me that spurious contact terms related to that part of the theory were not responsible for the two-loop problems.

Somewhere in the first half of 1970, I heard via Zumino that Faddeev (and Slavnov, as I learned later) [29] had established that already at the one-loop level the massless theory was not the limit of the massive theory. The difference was hiding in a symmetry factor for the one-loop ghost graphs: they have a factor of $\frac{1}{2}$ as compared to the Faddeev-Popov ghost loops of the massless case. In the summer of 1970, H. Van Dam and I reproduced and understood the argument and went further to consider gravitation [30]. Here we found one of the more astonishing facts in this domain: for gravity the limit from massive gravitons to zero mass is not the same as the massless theory (of Einstein). Thus a theory of gravitation with a massive spin 2 particle of exceedingly small mass (for example, of the order of an inverse galactic radius) would give a result for the bending of light by the sun that was discretely different (by a factor of $\frac{3}{4}$) from that of the massless theory. Thus by observing the bending of light in our solar system we can decide on the range of the gravitational field on a galactic scale and beyond. Many physicists (I may mention Kabir here) found this result hard to swallow. The discontinuity of the zero-mass limit as compared to the massless case has always been something contrary to physical intuition. Indeed, for photons there is no such effect. The work with Van Dam was actually my first exercise in the quantum theory of gravitation.

So, by the end of 1970, I was running out of options. I started to think of studying the difference between the massless and massive cases, more explicitly, to try to sort of subtract the massless theory from the massive theory, diagramwise. That would have produced a hint to the Higgs system. Indeed, the theory with a Higgs particle allows a continuous approach to the massless theory with, however, four extra particles. Furthermore, I was toying with the idea to see if the remaining infinities had a sign that would allow subtraction through some further interactions. Conceivably, all this could have resulted in the introduction of an extra particle, the Higgs, with interactions tuned to cancel unwanted divergencies, or to readjust the one-loop counter terms to be gauge invariant (in my paper the four-point counter terms were not). The result would have been in the worst kind of "Veltmanese" (a term used by Coleman to describe the style of 't Hooft's first article). However, that development never happened, and hindsight is always easy. That kind of work needed something else: a regularization procedure. Not only was the lack of a suitable method impeding further investigation or application of the results obtained so far to practical cases, but there was also the question of anomalies. It is at this point that 't Hooft entered into my program.

't Hooft became my student somewhere in the beginning of 1969. His first

task was writing what was called a "scriptie", a sort of predoctoral thesis, or *thèse troisième cycle* (in France). The subject was the anomaly and the σ model. That being finished in the course of 1969 he then started on his PhD work. At the same time he took part in my path-integral enterprise, so let me describe that.

I spent the academic year 1968–1969 at the University at Orsay, near Paris. During the summer of 1968 I was already there for the most part and met Mandelstam, who had been working on Yang-Mills theories as well. He had his own formalism [31], and we compared his results with mine. We did not note the notorious factor of 2 mentioned above: Mandelstam had studied the massless case, while my results applied to the massive theory. Boulware was also there. As a student of Schwinger he knew about functional integrals, and he later applied his skills to the subject [32]. It became clear to me that there was no escape: I had to learn path integrals. At the end of my stay at Rockefeller University somebody had already told me that there was work by Feynman [21] and also Faddeev on the massless theory. The article by Faddeev and Popov [33] was, as far as I was concerned, written in Volapuk. It also contained path integrals, and although I had accepted this article in my function as editor of Physics Letters, I had no inkling what it was about at the time (summer 1967). I accepted it then because of my respect for Faddeev's work. Just as well!

My method of learning about path integrals was lecturing on it, in Orsay. Ben Lee happened to be there as well, and he was also interested. With some diffculty I obtained the book of Feynman and Hibbs. (This was not easy; the students were busy making revolution and had no time for such frivolous things as path integrals. I thus sent around a note asking them to return the book prior to making revolution, which indeed produced a copy. This gave me the reputation of an arch reactionary, which I considered a distinction, coming as it did from Maoists.) Somewhere during these lectures a Polish physicist (Richard Kerner, now in Paris) produced another article by Faddeev, in Russian, and I asked him to translate it. I have never read that article, Ben Lee took it with him. I was simply not up to it and I still felt that I did not understand path integrals. So, returning to the Netherlands I decided to do it once more, and in collaboration with Nico Van Kampen we did set up a course in path integrals (autumn 1969). My then student 't Hooft was asked to produce lecture notes, which he did. I would say that then I started to understand path integrals, although I have never felt comfortable with them. I distrust them. 't Hooft had no such emotional ballast, and he became an expert in the subject. So, by the end of 1969 't Hooft had been educated in the σ model, anomalies, and path integrals.

't HOOFT

At this point 't Hooft showed unhappiness with the provisional subject that I had suggested, namely, the double-resonance peak of Maglic. He wanted to enter into the Yang-Mills arena. I then suggested that he investigate the mass-

less theory, with emphasis on finding a regulator method. This was so decided during a dinner, also attended by Van Kampen.

In studying the massless case 't Hooft used combinatorial methods (diagram manipulation) to establish various identities [34]. He could have used the Ward identities of my earlier paper, but I think he wanted to show that he could do better. Thus it came to pass that he never wrote the Slavnov-Taylor identities, an oversight that these two gentlemen quickly corrected [35,36]. 't Hooft derived mass shell identities, presumably enough for renormalization purposes.

Perhaps the main point that we argued about was the necessity of a gauge-invariant regularization scheme. He took the point of view that no matter what scheme one uses one simply adjusts the subtraction constants so that the Ward identities are satisfied, and that is all that is needed to renormalize the theory. Well, that is true provided there are no anomalies, and after some time he accepted the point. He developed a gauge-invariant method that worked up to one loop. A fifth dimension was used. Later, trying to go beyond one loop we developed the dimensional regularization scheme; in the summer of 1971 we had a rough understanding of that method. I should say that at all times I had an ulterior motive: I very much wanted an actually usable scheme. The existing methods (such as the Pauli-Villars scheme) are perhaps useful in doing quantum electrodynamics, but completely impractical for a Yang-Mills theory. I needed a good tool.

In an appendix to his paper 't Hooft presented, within the path-integral scheme, a gauge choosing method. I did not recognize this at all, but later, backtracking, I discovered that this was an evolved version of my original attempt at a change of gauge, including the "free field" technique. Russian physicists (Faddeev, Slavnov, Fradkin and Tyutin) took it over into the path-integral formalism, mangled, cleaned, and extended the method, mainly applying it to the massless case as well as (massless) gravitation. The actual scheme proposed by 't Hooft in the quoted appendix is the method that is mostly used today.

I am not going to describe the (substantial) Russian contributions here. This despite the fact that, like almost everywhere else, doing field theory was not very popular in the Soviet Union. I believe that in ref [1] a fair account has been given.

I am also skipping a description of the second paper of 't Hooft [37], introducing spontaneous symmetry breakdown and thus arriving at the renormalizable theories with massive vector bosons as known today. There are only two points that I would like to mention: I insisted that as much as possible the results should not depend on the path-integral formalism, i.e. that unitarity should be investigated separately, and secondly, that the issue of there being something in the vacuum not be made into a cornerstone. Indeed, once the Lagrangian including spontaneous symmetry breaking has been written down you do not have to know where it came from. That is how I wanted the paper to be formulated. I suspected that there might be trouble with this vacuum field, and I still think so, but that does not affect in any way

't Hooft's second paper. He sometimes formulates this as me opposing the cosmological constant, but at that time I did not know or realize that this had anything to do with the cosmological constant. That I realized for the first time during a seminar on gravitation at Orsay, at the beginning of 1974. See ref. [38].

So let me go on to the autumn of 1971. 't Hooft dived into massless Yang-Mills theory studying the issue of asymptotic freedom; I think that Symanzik put him on that track. I devoted much attention to the dimensional regularization scheme [39]. Again I refer the interested reader to ref. [1] for details, including the independent work of Bollini and Giambiagi.

After dimensional regularization was developed to an easily workable scheme I decided that it would be a good idea to write two papers: (i) a paper clearly showing how everything worked in an example, and (ii) a reasonably rigorous paper in which renormalizable gauge theories were given a sound basis, using diagram combinatorial techniques only. The result was two papers, one entitled "Example of a gauge field theory" [40], the other "Combinatorics of gauge field theories" [41]. The first was presented at the Marseille conference, summer 1972, where a preliminary version of the second paper was also presented. I have no idea how many physicists read the "Example" paper; I think it is a pity that we published it only in the conference proceedings and not in the regular literature. In that paper all one-loop infinities of the simple SU2 model with a two-parameter gauge choice were computed, and the informed reader may without any trouble use the counter Lagrangian given in that paper to deduce the β parameter for that theory (including a Higgs). That is the parameter relevant for asymptotic freedom. The calculations for this paper were fully automated and done by *Schoonschip*. When 't Hooft asked me about the divergencies of the massless theory as a check on his own calculations there was no problem doing that. I did not know about asymptotic freedom and did not understand the relevance of this particular calculation at that moment. He reported his result at the Marseille conference.

RADIATIVE CORRECTIONS

Much of my effort after 1972 were directed towards applying the theory, i.e., towards radiative corrections. In 1975 there was still considerable argument about neutral currents. Most people thought that the precise configuration contained in the Weinberg model was a must, not knowing that by choosing another Higgs sector one could adjust the Z_0 mass to any value. This was clearly a critical point, and Ross and I set out to investigate this issue [42]. This led to the introduction of a new parameter, now called the ρ parameter, that takes on the value of 1 for the simplest Higgs sector as chosen by Weinberg. The ρ parameter is essentially the square of the ratio of the charged vector-boson mass to the neutral vector-boson mass, with a correction related to weak mixing. This parameter has become an important part of today's physics, because it is the most sensitive location for radiative effects of

heavy particles, quarks or Higgs. At the Paris conference on neutral currents of 1974 or 1975 I presented a very short contribution, consisting of, I believe, only two transparencies. All I said was this: the neutral vectorboson mass can be anything. Here is a convenient way to parametrize that. To this day I am flabbergasted that nobody, but nobody at that conference seemed to have gotten the message. They kept on thinking that finding the precise quantitative amount of neutral-current effects as predicted by the Weinberg model (extended to quarks according to Glashow, Iliopoulos, and Maiani) was crucial to the applicability of gauge theories. In reality, had they found deviant results, the only consequence would have been a different Higgs sector.

In 1976 it became reasonably clear that the Standard Model including the simplest possible Higgs sector was the right model. Meaningful calculations on radiative corrections were now possible, and I set out to do them. It appeared that there were at least three families. The following issues were of immediate importance to me:

 (i) How many generations are there?
(ii) Is there an upper limit on the Higgs mass?

I will not enter into the argument on the number of neutrinos from astrophysics. Such arguments are less than airtight, because they build on the whole body of our understanding of the big bang and evolution of the Universe. Concerning the third generation, an interesting argument developed: what is the mass of the top quark? It would fill an amusing article to list all articles that made claims one way or the other, but I leave that to someone else. I realized that without a top quark the theory would be non-renormalizable, and therefore there ought to be observable effects becoming infinite as the top-quark mass goes to infinity. To my delight there was such a correction to the ρ parameter and furthermore it blows up proportional to the top-quark mass squared [43]. This is the first instance in particle physics of a radiative correction becoming larger as the mass of the virtual particles increases. That is our first window to the very-high-energy region. This radiative correction became experimentally better known and eventually produced a prediction for the top-quark mass of 175 GeV, which agrees with the result found when the top was discovered. This agreement also seems to indicate that there are no more generations, because there is little or no room for any quark mass differences in (hypothetical) new generations. Given the pattern of masses that we observe now, that appears unlikely, although it is strictly speaking not impossible.

From the beginning I was very interested in the Higgs sector of spontaneously broken theories. I started to look for a way to establish a limit on the Higgs mass; after all, if the Higgs is an essential ingredient of renormalization there must be terms in perturbation theory that cannot be renormalized away and that would be sensitive to the Higgs mass. It can easily be argued that the place to look for that is in the radiative corrections to the vector-boson masses, and the relevant parameter there is the ρ parameter, introduced in the paper with Ross mentioned before. As it happens, while there could have been an effect proportional to the square of the Higgs mass, it turned out

that that piece cancels out and only a logarithmic dependence remains [44]. This makes it very diffcult to estimate the Higgs mass on the basis of radiative corrections, and I have introduced the name "screening theorem" in this connection. Nature seems to have been careful in hiding the Higgs from actual observation. This and other facts have led me to believe that something else is going on than the Higgs sector as normally part of the Standard Model.

After that I started to set up a systematic scheme for the calculation of radiative corrections, together with Passarino. As he, together with Bardin, has written a book that has just come out I refer the interested reader to that book [45].

There was another motive in doing the radiative corrections. I wanted ultimately to compute the radiative corrections to W-pair production at LEP, because it was clear to me that those corrections would be sensitive to the Higgs mass. This then would suggest a value for the LEP energy: it should be high enough that radiative corrections to W-pair production were sufficiently large and could be studied experimentally. One would either find the Higgs or see important radiative corrections. That calculation was done together with Lemoine [46] and finished in 1980. I did not succeed in making the case sufficiently strong: no one understood the importance of such considerations at that time. Thus the LEP energy came out at 200 GeV, too low for that purpose. As the vector bosons were still to be found, few were prepared to think beyond that. Also, I have no idea if a 250- to 300-GeV LEP would have been possible from an engineering point of view, let alone financially.

PRESENT STATUS

The Higgs sector of the Standard Model is essentially untested. Customarily one uses the simplest possible Higgs system, one that gives rise to only one physical Higgs particle. With that choice the Z_0 mass is fixed to be equal to the charged W mass divided by $\cos(\theta)$, where θ is the weak mixing angle. Let us first establish here a simple fact: by choosing the appropriate Higgs sector one can ensure that the Z_0 mass is unconstrained. Furthermore the photon mass need not be zero and can be given any value.

In the early days a considerable amount of verbosity was used to bridge the gap between the introduction of the models in 1967 and the later-demonstrated renormalizability. Two of the terms frequently used to this day are "electroweak unification" and "spontaneous symmetry breakdown." As I consider these terms highly misleading, I would like to discuss them in some detail.

To what extent are weak and electromagnetic interactions unified? The symmetry used to describe both is SU2 × U1, and that already shows that there is really no unification at all. True unification, as in Maxwell's theory, leads to a reduction of parameters; for example, in Maxwell's theory the propagation velocities of magnetic and electric fields are the same, equal also to the speed of light. In the electroweak theory there is no such reduction of parameters: the mixing angle can be whatever, and that makes the electric coup-

ling constant $e = g\sin(\theta)$ a free parameter. If the Higgs sector is not specified, then the Z_0 mass and the photon mass are also free parameters. There is really no unification (apart from the fact that the isovector part of the photon is in the same multiplet as the vector bosons).

However, if one specifies the simplest possible Higgs system then the number of free parameters diminishes. The Z_0 mass is fixed if the weak mixing angle and the charged vector-boson mass are fixed, and the photon mass is necessarily zero. So here there seems to be some unification going on. It seems to me, however, utterly ridiculous to speak of "electroweak unification" when choosing the simplest possible Higgs system.

The question of spontaneous symmetry breakdown is more complicated. From my own perspective the situation is as follows. In 1968, I showed what I termed the one-loop renormalizability of the massive Yang-Mills theory. The precise meaning of that will become clear shortly. However, there is trouble at the two-loop level, so at the time I thought that there had to be some cutoff-mechanism that would control the (observable) divergencies occurring beyond the one-loop level. Actually, the Higgs system can be seen as such a cutoff mechanism. The Higgs mass becomes the cutoff parameter, and indeed this cutoff parameter is observable (which is the definition of non-renormalizability of the theory without a Higgs system). This parameter enters logarithmically in certain radiative corrections (the Z_0 mass, for example), and from the measurement of these corrections follows some rough estimate of the value of this parameter. However, part of the input is that the Higgs sector is the simplest possible; without that assumption there is no sensitivity at the one-loop level (because then the Z_0 mass is not known and the radiative correction becomes a renormalization of that mass). That is the meaning of one-loop renormalizable. Let us, however, assume that from a symmetry point of view things are as if the Higgs system were the simplest possible. Then from the radiative corrections the cutoff parameter (the Higgs mass) can be estimated.

From this point of view the question is to what extent we can be sure today that the cutoff system used by Nature is the Higgs system as advertised. Evidence would be if there is indeed a particle with a mass equal to the estimate found from the radiative corrections. But if there is none, that would simply mean that Nature uses some other scheme, to be investigated experimentally.

In all of this discussion the notion of spontaneous symmetry breakdown does not really enter. In the beginning this was a question that I kept on posing myself. Spontaneous symmetry breakdown usually implies a constant field in the vacuum. So I asked myself: is there any way one could observe the presence of such a field in the vacuum? This line of thought led me to the question of the cosmological constant [38]. Indeed, if nothing else, surely the gravitational interactions can see the presence of a field in the vacuum. And here we have the problem of the cosmological constant, as big a mystery today as 25 years ago. This hopefully also makes it clear that, with the introduction of spontaneous symmetry breakdown, the problem of the cosmological constant

enters a new phase. I have argued that a solution to this problem may be found in a reconsideration of the fundamental reality of space-time versus momentum space [47], but this is clearly not the place to discuss that. Also, the argument has so far not led to any tangible consequences.

So, while theoretically the use of spontaneous symmetry breakdown leads to renormalizable Lagrangians, the question of whether this is really what happens in Nature is entirely open.

CONCLUSION

The mind-wrenching transition of field theory in the sixties to present-day gauge-field theory is not really visible anymore, and is surely hard to understand for the present generation of field theorists. They might ask: why did it take so long? Perhaps the above provides some answer to that question.

REFERENCES

1. M. Veltman, "The Path to Renormalizability", invited talk at the Third International Symposium on the History of Particle Physics, June 24–27, 1992. Printed in "The Rise of the Standard Model", Lillian Hoddeson, Laurie Brown, Michael Riordan and Max Dresden Editors. Cambridge University Press 1997.
2. S.L. Glashow, *Nucl. Phys.* 22 (1961) 579.
3. S.L. Glashow, J. Iliopoulos and L. Maiani, *Phys. Rev.* D2 (1970) 1285.
4. S. Weinberg, *Phys. Rev. Let.* 19 (1967) 1264; A. Salam, in "Elementary Particle Theory", Stockholm 1968, N. Svartholm editor.
5. Y. Hara, *Phys. Rev.* 134 (1964) B701.
6. W. Bardeen, *Phys. Rev.* 184 (1969) 1848.
7. C. Bouchiat, J. Iliopoulos and Ph. Meyer, *Phys. Let.* 38B (1972) 519.
8. S.L. Adler, *Phys. Rev. Let.* 14 (1965) 1051.
9. W.I. Weisberger, *Phys. Rev. Let.* 14 (1965) 1047.
10. S. Weinberg, Nobel lecture 1979.
11. A. Salam, *Phys. Rev.* 127 (1962) 331.
12. M. Gell-Mann, *Physics* 1 (1964) 63.
13. M. Veltman, *Phys. Rev. Let.* 17 (1966) 553.
14. J.S. Bell, *Il Nuovo Cim.* 50A (1967) 129.
15. S.L. Adler, *Phys. Rev.* 137B (1965) 1022.
16. · M. Veltman, *Proc. Roy. Soc.* A301 (1967) 107.
    J.S. Bell and R. Jackiw, *Il Nuovo Cim.* A60 (1969) 47.
17. S.L. Adler, *Phys. Rev.* 177 (1969) 2426.
18. T.D. Lee and C.N. Yang, *Phys. Rev.* 128 (1962) 885.
19. T.D. Lee, *Phys. Rev.* 128 (1962) 899.
20. M. Veltman, *Physica* 29 (1963) 186.
21. R.P. Feynman, *Acta Phys. Pol.* 24 (1963) 697 (talk July 1962).
22. B.S. DeWitt, *Phys. Rev. Let.* 12 (1964) 742; Phys. Rev. 160 (1967) 1113; Phys. Rev. 162 (1967) 1195, 1239.
23. E.C.G. Stueckelberg, *Helv. Phys. Acta* 11 (1938) 299. See ref. [1] for other references to this subject.
24. M. Veltman, *Nucl. Phys.* B7 (1968) 637. See also M. Veltman, Copenhagen lectures, July 1968. Reprinted in "Gauge theory – past and future", R. Akhoury, B. De Wit, P. Van Nieuwenhuizen and H. Veltman, Editors. 1992, World Scientific, Singapore, p 293.
25. S.L. Glashow and J. Iliopoulos, *Phys. Rev.* D3 (1971) 1043.
26. M. Veltman, *Nucl. Phys.* B21 (1970) 288 [16 Apr 1970].
27. M. Veltman, Proc. Topical Conf. on Weak Interactions, CERN, Geneva, 14–17 Jan 1969, CERN yellow report 69-7, page 391.

28. J. Reiff and M. Veltman, *Nucl. Phys.* B13 (1969) 545.
29. A.A. Slavnov and L.D. Faddeev, *Teor. Mat. Fiz.* 3 (1970) 18.
30. H. Van Dam and M. Veltman, *Nucl. Phys.* B22 (1970) 397.
31. S. Mandelstam, *Phys. Rev.* 175 (1968) 1580.
32. D. Boulware, *Annals of Phys.* 56 (1970) 140.
33. L.D. Faddeev and V.N. Popov, *Phys. Let.* 25B (1967) 29.
34. G. 't Hooft, *Nucl. Phys.* B33 (1971) 173.
35. A.A. Slavnov, *Theoretical and Mathematical physics* 10 (1972) 153 (English translation page 99).
36. J.C. Taylor, *Nucl. Phys.* B33 (1971) 436.
37. G. 't Hooft, *Nucl. Phys.* B35 (1971) 167.
38. M. Veltman, Cosmology and the Higgs mechanism. Rockefeller University preprint May 1974. M. Veltman, *Phys. Rev. Let.* 34 (1975) 777.
39. G. 't Hooft and M. Veltman, *Nucl. Phys.* B44 (1972) 189.
40. Renormalization of Yang-Mills fields and applications to particle physics, Marseille Conference June 19-23, 1972. Proceedings edited by C.P. Korthals-Altes, p 37.
41. G. 't Hooft and M. Veltman, *Nucl. Phys.* B50 (1972) 318.
42. D.A. Ross and M. Veltman, *Nucl. Phys.* B95 (1975) 135.
43. M. Veltman, *Nucl. Phys.* B123 (1977) 89.
44. M. Veltman, *Acta Phys.* Pol. B8 (1977) 475.
45. D. Bardin and G. Passarino, The Standard Model in the Making. Clarendon Press, Oxford 1999.
46. M. Lemoine and M. Veltman, *Nucl. Phys* B164 (1980) 445.
47. M. Veltman, *Acta Phys.* Pol. B25 (1994) 1399.