

Shelter Island II

Proceedings of the 1983 Shelter Island Conference on
Quantum Field Theory and the Fundamental Problems of
Physics

*edited by Roman Jackiw, Nicola N. Khuri, Steven Weinberg,
and Edward Witten*

The MIT Press
Cambridge, Massachusetts
London, England

It is a great honor to be called on to speak at the beginning of this meeting of very distinguished colleagues, celebrating the thirty-sixth anniversary of that remarkable gathering here at Shelter Island. I hope you'll forgive my poorly organized remarks and especially the randomness of my references: I shall attribute research to the authors only occasionally (and no doubt wrongly), but the serious technical lectures that follow will take care of the errors and omissions, in both the physics and the references.

Thirty-six years ago I was just starting to learn particle physics, and it struck me very forcibly that theoreticians were in disgrace. They were in trouble on two grounds: the experimental muon didn't agree with the theoretical meson in its properties; and all radiative corrections in the one theory available—quantum electrodynamics—gave infinity. Just at that time, in the year of Shelter Island, things righted themselves. Marshak's two-meson hypothesis, followed by the experimental discovery of the pion, fixed one difficulty. And the measurement and calculation of the Lamb shift fixed the other one.

Since then, we theoreticians have been riding high. Every once in a while, of course, experiments give us a jolt, as in the case of *CP* violation or the existence of the third family of spin $\frac{1}{2}$ fermions. (And of course we never knew where the second one came from, or the first one either, for that matter!)

In the meantime—in fact, a decade ago—we theoreticians in partnership with our experimental colleagues have largely solved the problems that confronted us thirty-six years ago, while replacing them with others. We seem to have a renormalizable theory of strong, electromagnetic, and weak interactions that works. In this talk I shall assume that it is correct, with perhaps minor modifications, at least for energies below 1 TeV. It is of course a Yang-Mills theory, based on color $SU(3)$ and electroweak $SU(2) \times U(1)$, with 3 families of spin $\frac{1}{2}$ leptons and quarks, their antiparticles, and some spinless Higgs bosons in doublets and antidoublets of the weak isotopic spin to break the electroweak group down to U_1 of electromagnetism. I shall assume also that the *t* quark will be discovered by our experimental friends to complete the third family. The Higgs dynamics is, of all this, the part least tested by experiment and the most artificial looking. Otherwise the theory has considerable confirmation, especially now that the intermediate boson, X^\pm (which some people still call W^\pm , although I'm sure that will stop after a while) has been found, twenty-five years after we predicted it.

Essentially, QED has been generalized. QED was invented around 1929 and has never changed; but in the years after the Shelter Island meeting it was understood to be renormalizable, to give definite predictions to every order in perturbation theory. Now, QED has merely been generalized to include the strong and weak interactions along with electromagnetism, the quarks and neutrinos along with the electrons, and the mysterious 3 families. The new theory has similar characteristics, but it has many dimensionless parameters instead of one—very many, especially in connection with the Higgs bosons, which couple to themselves, and then couple separately to all the fermions, with coupling constants adjusted to give all the fermion masses.

The renormalizability means that we can absorb all the infinities into these parameters and into the mass scale, but it doesn't help us to calculate. This is a generalization of the uncalculability of $1/137$ in QED. Now, that was always an issue, but people hesitated to take it up in the old days. (I think that was largely because they didn't want to get mixed up with Eddington and his "intricate" and "extriculate," and thus be subjected to ridicule.)

When I say "renormalizability" I mean perturbative renormalizability. After thirty-five years of trying, theorists have been unable to invent any scheme for rendering finite or renormalizable a theory that isn't so in perturbation theory. Now, of course some day someone may solve that problem, but I shall assume today that finiteness and renormalizability apply to perturbation theory. I remember many occasions when it had been announced that somebody had solved the problem. The day I arrived at the Institute for Advanced Study as a very young man, I was told that Ning Hu had shown the pseudovector meson theory to be nonperturbatively renormalizable, and I've been told the same kind of thing many, many times since, just the way mathematicians are told all the time that somebody has proved Riemann's conjecture.

(I heard last night from André Martin that Hilbert sent a telegram before he gave a speech in Berlin, saying that he had proved Riemann's conjecture. Then he arrived and the audience was very disappointed in the speech, because he said nothing about such a proof. It turned out he had sent the telegram in case he was killed on the way, on his first airplane trip, of which he was terrified. Disappointed as you may be in my remarks today, it probably won't be as bad as that.)

As usual, solving the problems of one era has shown up the critical questions of the next era. The very first ones that come to mind, looking at the standard theory of today, are

- Why this particular structure for the families? In particular, why flavor-chiral, with the left- and right-handed particles being treated differently, rather than, say, vectorlike, in which left and right are transformable into being treated the same?
- Why 3 families? That's a generalization of Rabi's famous question about the muon, which I'll never forget: "Who ordered that?" The astrophysicists don't want us to have more than 3 families. Maybe they would tolerate a 4th, but no more, with massless or nearly massless neutrinos; it would upset them in their calculations of the hydrogen and helium isotope abundances. Of course, if the neutrinos suddenly jumped to some huge mass in going from known families to a new one, then they would be less upset.
- How many sets of Higgs bosons are there in the standard theory? Well, the Pececi-Quinn symmetry, which I'll mention later, requires at least 2, if you believe in that approach. If there's a family symmetry group, there may be more, because we may want a representation of the family symmetry group: maybe there are 6 sets of 4 Higgs bosons; nobody knows.
- Why $SU(3) \times SU(2) \times U(1)$ in the first place? Here, of course, there have been suggestions. We note that the trace of the charge is zero in each family, and that suggests unification with a simple Yang-Mills group at some high energy, or at least a product of simple groups with no arbitrary $U(1)$ factors. If the group is simple or a product of identical simple factors, then we can have a single Yang-Mills coupling constant.

I should like to emphasize in this talk the thread of renormalizability, finiteness, and calculability. It runs all through our work from Shelter Island I to Shelter Island II—through QED, $SU(3) \times SU(2) \times U(1)$, unified Yang-Mills theories, attempts to unify with gravity, and superstrings. Hence the title: "From Renormalizability to Calculability?"

After the triumphs of thirty-five years ago, many theoreticians—although not my colleague Richard Feynman—changed their minds about the significance of renormalizability. In QED, after all, e and m are arbitrary. Only one of them is dimensionless (and as I said before, nobody wanted to get mixed up with Eddington and his calculation of alpha). So it seemed perfectly all right just to say that the renormalized theory was *the*

theory and that we must put in the renormalized e and m . Feynman persisted in regarding renormalization in a different way, as it's treated in the cartoon, courtesy of Pierre Ramond and the artist Joan Cartier.*

Now—is it really like that? Well, in the 1970s I would say that Richard Feynman was led into error by this idea in the sense that he disbelieved charm and neutral weak currents on the grounds that they were needed for a renormalizable theory, and that he thought the requirement of a renormalizable theory was not a good criterion. However, at a deeper level he may still be right. Many theorists are now raising the question of whether our renormalizable broken Yang-Mills theories [the $SU(3) \times SU(2) \times U(1)$ and some higher unified one] are not, after all, just renormalizable phenomenological theories in which the cutoff dependence is replaced by the existence of a huge number of lumped parameters—renormalized dimensionless parameters calculable only in a more fundamental theory of a different kind that would really be finite.

Right in the $SU(3) \times SU(2) \times U(1)$ theory we're faced with the situation that quantum chromodynamics has a hidden uncontrollable parameter, the mixing angle between chromoelectric and chromomagnetic fields that violates CP unless it can be rotated away by a γ_5 transformation on some flavor of quark field. Well, if the u quark actually had 0 ultraviolet mass, that would be OK—but it seems to have roughly half the ultraviolet mass of the d quark rather than 0 times it. We seem to have three more choices: we dial the angle to 0 and look for a theory where the radiative corrections are somehow finite and small enough to be compatible with the experimental levels of CP violation; or we put in a cutoff and say that with the cutoff the corrections are small; or else we find a symmetry that would rotate the angle away. This is the Peccei-Quinn symmetry, which involves at least doubling the Higgs boson; the associated $U(1)$ symmetry gives us a light modified Goldstone Higgs boson called a higglet, with a very small mass. In connection with the higglet, what happens is that the product of the coupling constant of the higglet to each fermion times the vacuum expected value of some Higgs field is fixed, and if you make one very large, the other one becomes small. It has been suggested in the last few years that maybe the expected value is very large and each coupling constant very tiny, and that

* Editors' note: The cartoon referred to shows a cleaning woman tidying up the office of the Field Theory Group by sweeping infinities under the rug; the caption has the cleaning woman saying that she's "got this one renormalized."

then you would get an invisible higglet that no experimental or astrophysical test yet devised could find. Also it has been claimed that if you make the expected value too big, and the coupling constants too small, then the universe is somehow unhappy, so that one must have an expected value somewhere around 10^{12} GeV. You may ask whether that is an important energy in some other connection, and actually it has turned up in some recent other research as a possibly fundamental energy scale. By the way, some people have called the higglet by another name, in which case it's extremely easy to discover in any supermarket.*

Here, as elsewhere, we seem to have to dial various renormalized quantities to small values. The situation of having numerous arbitrary dimensionless parameters is even more humiliating when some of them are very small. Now there's a hierarchy of remedies for this. (If you want to use that word—it means "sacred rule" and has to do with priestly government. I don't know what it's doing in our subject!) First of all, we'd like to interpret small or nearly symmetric quantities as coming from a slightly broken symmetry; otherwise they don't make any sense to us. The small quantities would approach zero in the limit of perfect symmetry. Second, becoming more ambitious, we would like them if dialed to small values to stay there and not acquire uncontrollable radiative corrections resulting in unknown values, which could again be large. Third, and even better, we would like to be able to calculate them and know why they are small. We should distinguish among these three objectives. When we can't avoid dialing some renormalized quantity to a small value (knowing that if the unrenormalized quantity were small there would be infinite and therefore uncontrollable corrections reparable only by renormalization), that situation has recently been described as a problem of "naturalness." In practice each theoretician seems to dial one or more quantities to a small value in his own work and then attacks other theorists for performing unnatural acts when they do the same thing. In fact, the idea of avoiding dialing renormalized quantities to symmetric or nearly symmetric values is really a very old one. Many of us have thought for thirty years (and said) that simplicity and symmetry lay typically in the direction of higher energies, and that effective coupling constants and effective masses as followed upward in energy via the renormalization group would show their true underlying symmetries in the limit of high energy. We showed thirty years ago that symmetry in the limit

* Editors' note: At this point a box of Axion laundry presoak is held aloft by Gell-Mann.

of high energy was essentially equivalent to symmetry applying to unrenormalized quantities. We indicated then that we should not dial renormalized quantities to symmetric or nearly symmetric values, or to zero. Now, in today's standard Yang-Mills theory, or its unified generalization, because the broken symmetries are violated by the expected values of Higgs fields, which give ultraviolet fermion masses and vector boson masses, the symmetries do show up as you go to energies large compared with those masses.

It has been suggested in the last few years (and not followed up very much) that there may be other kinds of symmetry. For example, authors including Iliopoulos, Tomaras, and Maiani have shown that coupling constants obeying the renormalization group equations, as in other problems in nonlinear systems dynamics, can tend toward symmetry merely by virtue of the properties of the equations. That can happen if the energy is reduced, and such symmetries would never be fully realized. The equations would try to approach a symmetric situation as energy approaches zero, but they would encounter the symmetry-violating masses and the symmetries would never achieve full expression. However, I shall not pursue this fascinating heresy. If we use the usual description of broken symmetry, then what we would like best when we see a small or nearly symmetric renormalized quantity is to have it reflect a broken symmetry, with a finite, calculable correction that will explain the small number. This is actually not easy to do in our broken Yang-Mills theories.

Now all of you know that it's tempting to try to include the $SU(2) \times U(1) \times SU(3)$ theory in a unified Yang-Mills theory, and that what happens if you follow the three coupling constants of the three groups up in energy by the renormalization group equations, assuming fermion families like the ones we see today, is that the coupling constants approach one another at around 10^{15} GeV. The fact that all three come together is nontrivial and suggests that maybe there really is some kind of unification with, say, a simple Yang-Mills group that is badly broken. If you have just 3 fermion families, then the value of the coupling constant near the unification energy is something like $1/40$, and that would replace $1/137$ as the dimensionless parameter that requires explanation.

The simplest version of this kind of theory uses $SU(5)$, as you know, and employs a large breaking, which smashes the group down to $SU(3) \times U(1) \times SU(2)$ at around 10^{15} GeV by means of one or more adjoint representations of higgses, and a weak breaking based initially on 5 's and $\bar{5}$'s. Unfortunately such a weak breaking presents some difficulties; while the

masses of the tau lepton and the b quark have the right ratio when the finite renormalizations are taken into account, the ratio of the muon and s quark masses (if we understand the ultraviolet mass of the s quark, which I happen to think we do) doesn't come out so good. You don't even try to get agreement in the case of the electron and the d quark; you assume that their masses are so small that they must arise from some other mechanism. The discrepancy can be corrected if we complicate the theory by introducing μ - s additional higgses belonging to other representations of $SU(5)$.

Another possible unified theory makes use of $SO(10)$ and puts each family of fermions into a 16, which on reduction to $SU(5)$ gives a 5 and a $\bar{10}$ and a singlet that is an inactive left-handed antineutrino. A further generalization uses the graph E_6 , with each fermion family in a 27.

So ... what I'll call "quantum unified dynamics" (some people call it "tripe" or "guts" or something of that kind), while it has very much to recommend it, is still complicated; and it still doesn't explain the fermion representations and multiplicities or the very large number of coupling constants; it has some of the same difficulties as the $SU(3) \times SU(2) \times U(1)$ theory. The word "guts" is appropriate in this discussion in that it takes enormous guts to suppose that we can extrapolate our ideas from 100 GeV, where we are experimentally, to 10^{15} GeV. Fortunately, there are consequences of these theories that don't involve experiments at 10^{15} GeV.

It should be mentioned that 10^{15} GeV is so close to the Planck mass (2×10^{19} GeV, where the strength of quantum gravity is around unity) that we may consider two possibilities:

One is to regard that ratio as being very important, so that there are two stages—a stage of unification with gravity around the Planck mass, and a stage where gravity can be ignored and Yang-Mills theories unified.

The other is to suppose that there is only one stage and that 10^{15} is just another name for 10^{19} , so to speak.

I would say we don't really know which of those to believe, although the former certainly looks tempting.

One famous important feature of quantum unified dynamics is that without special measures one tends to get proton decay at a rate comparable to the experimental limits. That is one probe of very high energy phenomena at low energies. Another possible probe is connected with neutrino masses, as we shall discuss shortly. It may be that in the future we shall find some other such tests or probes at moderate energies attainable by experiment, say 1 TeV or so. That would be very useful, and also provide additional justification for constructing our next machines. Astrophysical

and cosmological tests are also of some value. Basically, what we're doing is looking for rules that obtain in $SU(3) \times SU(2) \times U(1)$, but are violated in a unified theory in a detectable manner.

Anyway, Sakharov's ingenious suggestion many years ago that proton decay, a suitable CP violation, and nonequilibrium conditions in the very early universe could explain the predominance of matter over antimatter was sharpened by many clever theorists after quantum unified dynamics suggested that the proton decay be taken seriously. And it looks now like an idea that we would hate to lose, especially since a rough prediction of the number of baryons over the number of photons can be given, and is not that far off. So quantum unified dynamics that has proton decay is desirable; and although scenarios are available for complicating the theory to avoid proton decay, we should probably not invoke them. An analogous scenario is found naturally in the simplest version of $SU(5)$ and gives rise to conservation of baryon number minus lepton number. As you know, that restricts proton decay so that it can yield a positron plus mesons but not an electron. It also has another effect; neutrino masses are made impossible. That result depends, though, on the group and on the choice of transformation properties of the particular Higgs field components that are nonzero in the vacuum. If you go to $SO(10)$, which is a rather nice generalization of $SU(5)$, not only do you add to each fermion family a right-handed neutrino and a left-handed antineutrino, but also you can easily arrange the representations so that you violate the conservation of the number of baryons minus the number of leptons.

Neutrino masses then arise in the following ways: you can get a huge Majorana mass m_M for the right-handed neutrino; you can have a normal-size Dirac mass m_D connecting the left- and right-handed neutrinos; and to order m_D^2/m_M you can get a small left-handed neutrino mass as well. Now, these left-handed neutrino masses are quite interesting astrophysically and cosmologically, because we need dark matter to bind the galaxies and clusters gravitationally. Dark matter is important to the universe as well; here not enough is known about galactic evolution for astronomers to be sure, but it looks as if enough dark matter is needed to make the universe at least approximately flat asymptotically. One possibility for that dark matter is for it to consist of elementary particles, and one possibility for the elementary particles is to have neutrinos with tiny masses. The sum of those left-handed neutrino masses would have to be of the order of tens of eV if all the dark matter consists of neutrinos. Larger masses for the

neutrinos would give trouble: the universe would be overclosed (except in the case of enormous masses, which would lead to a totally different regime). If there are masses for these left-handed neutrinos, then there will also be comparable or somewhat smaller transition masses among them, and those are being actively sought in the laboratory. Other candidate particles for dark matter are now envisaged by theoreticians, including higgslets.

Let us return now to the question of mysterious small numbers. In unified Yang-Mills theory, we have a number of them and they are very small. The renormalization-group-invariant mass, which is sometimes called Λ_{QCD} , in units of the unification mass, gives something like 10^{-16} . The mass of the weak bosons, which comes from e times a Higgs expected value, in units of the unification mass, comes out around 10^{-13} . The masses of the various fermions, which come from g coupling constants multiplied by expected values of the various Higgs φ 's, where those g 's are Higgs couplings to the fermions — range from 10^{-13} for the t quark (if they find it around present energies), down to 10^{-18} for the electron. We want to know why these are so small, and also why they are of the same general order of smallness: after all, one of them could be $10^{-1.001}$! In the case of $\Lambda_{\text{QCD}}/m_{\text{UNIF}}$, we know that it comes from the logarithmic change of coupling constants as we move up from Λ_{QCD} to the unification mass, and so it can be represented as $\exp(-C/\alpha_{\text{UNIF}})$, where C is a known number and α_{UNIF} is the unified fine structure constant. We are then led to the question of why the fine structure constant is small compared with 1 — but at least we don't have to explain an exponential smallness any more! What about the other small constants? Again, we can be slightly ambitious, moderately ambitious, or very ambitious. We can try to fix them so that if we dial them to small values, they stay there; or we can try to explain why they're nearly zero — for example, by showing that they also correspond to such exponentials (in certain models that's true); or we can actually calculate all of the small numbers, and that has so far eluded everybody. But the last is what we really want: to be able to calculate the constants. To express some of them as $\exp(-C/\alpha)$ is OK, but then we ought to be able to calculate the coupling constant involved.

Now, in $SU(3) \times SU(2) \times U(1)$ Yang-Mills theory, and especially in unified Yang-Mills theory, a great deal of effort has gone into trying to make modest improvements in this "naturalness" situation. In quantum unified dynamics, as far as I know, no one has ever given a thorough explanation, free of difficulties, of the small ratios I've just listed — why

they're all small and why they're all of the same general order of smallness. Certainly no one has calculated all of them!

However, in some schemes there has been partial success, especially in showing that if the constants are dialed to small values they stay small. The greatest difficulties of "naturalness" occur in connection with keeping the masses associated with the Higgs bosons low, as required, for example, to keep the weak breaking tiny in quantum unified dynamics (the ratio of 10^{-13} or so). We can regard the standard $SU(3) \times SU(2) \times U(1)$ theory as a phenomenological renormalizable theory within quantum unified dynamics—if there is a quantum unified dynamics. The well-known tendency of spinless bosons in field theory to get quadratic self-masses emerges here as a difficulty for the weak breaking higgses to avoid acquiring large masses from the cutoff, so to speak, when we consider the small theory as being a renormalizable effective field theory within the big one. Most of the nostrums of the last decade—fashions that last a year or two, typically—have been connected with fixing certain Higgs boson masses to be low, first tying them down to zero, and then, with corrections, to small values more or less calculable depending on the nostrum. One type of prescription had to do with discrete symmetries. More recent attempts can be classified in two ways: according to the degree of success in explaining the small parameters (success that's never been complete, of course) and according to which "elementary" particles are treated as composite, if any.

The proper name, I would suggest, for an elementary field in quantum field theory is "haplon." I encourage every one to use that name. It comes from the Greek for "single" or "simple." In fact it's cognate in Indo-European to "simple" because the Indo-European "s" becomes an "h" in Greek or Welsh. Sun, for example, is "helios" in Greek and "haul" in Welsh. The "haploid" generation is very familiar in biology, implying a single set of chromosomes instead of two, and so forth. This choice of name has already received the approval of many physicists—Dimopoulos, Nanopoulos, and Iliopoulos, and for the benefit of my French friends I add Rastopopoulos.

The question is, which of our familiar "elementary" particles are kept on as haptons? And which ones are sacrificed to complexity? Everyone's favorite candidate for being thrown to the wolves is the higgs. One idea was to fabricate them out of new fermions and antifermions possessing a new, exactly conserved color called "supercolor," "primed color," "hypercolor," "technicolor," or some such, with a Λ' (renormalization-group-invariant mass) much bigger than Λ_{QCD} . Quarks and leptons and the new

fermions would all lack ultraviolet masses. They would have only infrared masses, which would look ultraviolet to us at energies much smaller than Λ' . This scheme hasn't been too popular lately, perhaps because a generalization of the primed color group to a larger group, approximately conserved, that would explain the concentration of mass in the highest family of quarks and leptons, gave rise to excessively large neutral current interactions connecting one family with another.

Some schemes have made the quarks and leptons composite. In other schemes, the X^\pm bosons, which some people call W^\pm , and the Z^0 are easily sacrificed, since they are not massless gauge bosons for exact symmetries. The final step would be to make the photon and gluons composite, so that the whole $SU(3) \times SU(2) \times U(1)$ theory would contain only phenomenological fields, all composite objects. In the case of photon and gluons, one should check in any particular theory whether their masslessness and the exact conservation of charge and color are preserved when they are composite.

The favorite method nowadays for gaining control of corrections to small quantities is to apply $N = 1$ supersymmetry either to the standard theory or to quantum unified dynamics; as you know, $N = 1$ supersymmetry connects each state of a given J_z to one with a value of J_z differing by a half-unit. The spin 0 higgses are thus connected to otherwise unwanted spin $\frac{1}{2}$ superpartners, which one can tie down near zero mass by using approximate chiral invariance. In this way (or in other related ways) one can arrange for the light higgses to stay light. Their spin $\frac{1}{2}$ superpartners, like other unwanted superpartners, can then be pushed up to higher masses where they would not have been detected.

There is a whole "slanguage" that has evolved recently to describe new particles required by supersymmetry. I must admit to having been present when Glennys Farrar and Pierre Fayet coined names like "photino" and "gluino," for the spin $\frac{1}{2}$ partners of the spin 1 bosons, and "goldstino." The "goldstino" is the spin $\frac{1}{2}$ Goldstone fermion with zero mass that we need in order to allow spontaneous breaking of supersymmetry; and supersymmetry, if it is an exact symmetry, has to be broken spontaneously because we do not see a universal degeneracy of fermions and bosons. In supergravity, which we will discuss soon, there is a gravitino (spin $\frac{3}{2}$) that accompanies the graviton (spin 2). The gravitino is ideally poised for eating the goldstino and thereby acquiring mass, and undoubtedly that's what happens if supersymmetry is right. The goldstino can be either elementary or composite: it can

occur in the list of haptlons or it can arise dynamically; either way it is there for the hungry graviton to eat. The rest of the slanguage is needed because the spin $\frac{1}{2}$ quarks and leptons are accompanied by unwanted spin 0 partners called squarks and sleptons, and the higgsos, as we indicated before, have unwanted spin $\frac{1}{2}$ partners called, I suppose, shiggsos. (It sounds awful, but some people can say all this with a straight face!) All of these are accompanied by new spin $\frac{1}{2}$ and spin 0 particles, superpartners of each other, which nobody wants otherwise, but which are there in order to break the supersymmetry in the manner prescribed by O'Raifeartaigh. (His name, by the way, is written in a simplified manner; the "f" should really be "fthb.")

There has been some success along these lines, but more in the way of quantities staying small when dialed to be small than in the way of small quantities genuinely explained. Meanwhile, experimentalists are delighted that at least some theoreticians have made the desert bloom, and in fact turned it into a jungle.

Supertheories tend to have reduced divergences, and in some cases are completely finite in perturbation theory. Most applications exploit those properties, but one recent area of research ignores them, in what turns out to be a very interesting way. That is to say, some theorists take $N = 1$ supergravity, with one graviton and one gravitino, and couple it to the $N = 1$ super-Yang-Mills super-matter that we've just been describing. Now this is a crime, from the point of view we are adopting, because when gravity is accompanied by external matter, or when supergravity is accompanied by external supermatter, the divergences in perturbation theory are severe. However, one introduces a cutoff around the Planck mass, and then one calculates. New quantities of order unity occur, but one can only guess their values, so there is a good deal of freedom. One notices that a lot of familiar, important calculations are considerably altered if these numbers are correctly estimated. For example, the rate of proton decay and the preferred decay modes are changed (avoiding trouble with present experiments, by the way). Small neutrino masses, if they exist, can be altered. Quark-to-lepton mass ratios, which give some difficulties, as we indicated, can be improved if the unknown quantities come out right, and so on. Here supergravity is essentially a wild card that enables us to play the game in a much more exciting manner. However, since the quantities can only be estimated rather than calculated, it is a bit dangerous.

To return to our principal theme, what we want is, of course, a fundamental underlying theory. Whether the haptlons include all, some, or none

of our familiar elementary particles, except perhaps the graviton, we want everything to come out finite with no dimensionless parameters to start with, or at most one dimensionless parameter that gets renormalized, and a rest from the adjustability of numerous dimensionless quantities, of which, I think most of us are tired. Whether our impatience is justified, we don't know; it may be hundreds of years before we can get such a theory, if ever, but it's certainly what we would like.

West will, I think, tell us about $N = 4$ super-Yang-Mills and related theories that exhibit finiteness (no infinite charge renormalization any way),* but they have some other problems, such as that all particles have to be put in the the adjoint representation of the gauge group, which is conceivable but not terribly nice unless the graph is exceptional. Getting scale-invariance violation off the ground is also rather difficult in those theories, but still they are interesting.

It is also possible that there are ordinary $N = 1$ super-Yang-Mills theories that are finite; if there are any that agree with phenomenology, that would be exciting.

But the best candidates we have for fundamental theories in which there is unification of all interactions and all particles, at most one dimensionless parameter, and possible renormalizability or finiteness are theories related to $N = 8$ supergravity, where we have 8 supersymmetries, 8 gravitinos, and so forth. In fact, we have a list of haptlons with 1 graviton, 8 gravitinos, 28 spin 1 bosons, 56 spin $\frac{1}{2}$ fermions, 70 spin 0 bosons, and a partridge in a pear tree. In this theory, the number of dimensionless parameters is zero or one. There is κ , the square root of Newton's constant, which is dimensional; if that's all there is, then the theory is characterized by a chiral $SU(8)$ symmetry. However, there is also the option of introducing a dimensionless self-coupling parameter e , which gives a Yang-Mills character to those 28 spin 1 bosons; they gauge $SO(8)$, and the symmetry is thereby reduced from $SU(8)$ to $SO(8)$. This theory is only mildly divergent, if at all. In fact, ordinary gravity without external matter is only mildly divergent if at all! Gravity is suspected (but has not been found guilty) of a logarithmic divergence in two loops. The various supergravities are suspected (but have not been found guilty) of logarithmic divergences at higher numbers of loops. For $N = 8$ supergravity, there might be a divergence at three loops; some investigators think there will be no trouble before seven loops; and perhaps there is no trouble at all. There is no room for external supermatter

* Editors' note: The paper read to the conference by West appears in this volume.

in connection with $N = 8$ supergravity. All haptlons are in the same supermultiplet.

There are variations of $N = 8$ supergravity. One is $N = 7$ supergravity, which has the same list of particles and a slightly reduced symmetry. Then there are generalizations to higher numbers of spatial dimensions, which are extraordinarily interesting, and, on restriction to four dimensions, permit reductions of symmetry and the generation of coupling constants. And finally there are superstring theories, which are the most elegant candidates for a unified description of Nature.

First, I would like to describe the most serious problem with this whole set of theories, $N = 8$ supergravity and all the beautiful modifications thereof; namely, how to relate it to what we know. The work on Yang-Mills and super-Yang-Mills theories that I described was unsatisfactory in some respects, but at least it dealt in part with familiar objects, and we were able to relate all of it in some way or other to experiment. Here we go into a different realm where everything is very hopeful, very beautiful, exceedingly promising, probably renormalizable or maybe even completely finite, free of parameters—but with no obvious connection with experiment! We have somehow to make a bridge between these two domains.

Yuval Ne'eman and I looked at this question a long time ago, before supergravity was written down, but after $N = 8$ supergravity could be envisaged. We studied the quantum numbers involved; in particular, we assumed that the $SO(8)$ would be gauged, and we noticed, obviously, that $SO(8)$ is too small to include $SU(3) \times SU(2) \times U(1)$. Well, we tossed out the $SU(2)$ bosons—they're expendable, as we said, because they are not massless and their symmetries are violated in some mysterious manner by higgsos. That leaves $SU(3) \times U(1)$, which we tried to identify with color and electric charge. We found that the spin $\frac{1}{2}$ fermions, while they look almost right, don't come out quite the way we want. The fundamental **8**, in which the gravitino appears, would be a triplet and an antitriplet of color and two singlets, with charges q , $-q$, q' , and $-q'$, respectively. The easiest choice to take is $q' = 0$ and $q = -\frac{1}{3}$; then the fermions have some integral and some fractional charges, so there are things that look like leptons and things that look like quarks. That is very attractive, but one also gets a **6**, a $\bar{\mathbf{6}}$, and an **8** of color, which we don't see, and too few quarks and leptons. The total number of spin $\frac{1}{2}$ fermions is all right, but some are wasted on sextets and octets.

Now let me show you a curiosity. I apologize for wasting your time on

this scheme, which has probably nothing to do with physics; I offer it in the hope that somebody here can figure out something to do with it, because I can't. It is a last-ditch effort to salvage the identification of the spin $\frac{1}{2}$ haptlons of $N = 8$ supergravity with the fermions that we know. There are the right number, in a certain sense. If we assume that the goldstinos are haptlons and take 8 of them out of the 56, then we are left with 48, which is 3 times 16, and, as we mentioned in connection with $SO(10)$, there may be 16 left-handed fermions in each family if an inactive left-handed antineutrino is included. The symmetry is reduced from $SO(8)$, to $SO(7)$, breaking **56** into **48** + **8**. Of course $SO(7)$ contains $SU(3) \times U(1)$. We introduce values of the $U(1)$ generators that permit the same assignments for the 8 goldstinos and the 8 gravitinos, and then we examine the $SU(3) \times U(1)$ assignments of the 48 spin $\frac{1}{2}$ fermions. We find **(8 + 1, $\frac{1}{2}$)**; **(6 + 3, $-\frac{1}{6}$)**; **(3, $-\frac{1}{6}$)**; and **(3, $-\frac{5}{6}$)**. These correspond to the observed quarks and leptons if we assign the charge $\frac{2}{3}$ quarks and the neutrinos to a **3** of family $SU(3)$; we assign the charge $-\frac{1}{3}$ quarks and the negatively charged leptons to a **3** of family $SU(3)$; we identify the $U(1)$ generator with electric charge minus $\frac{1}{6}$ for the $\bar{\mathbf{3}}$ of family $SU(3)$ and with the electric charge plus $\frac{1}{6}$ for the **3** of family $SU(3)$; and we identify the $SU(3)$ of the theory with the diagonal $SU(3)$ of color $SU(3)$ times family $SU(3)$. These are a great many ifs and I have no idea how to justify them. Under family $SU(3)$, the charged weak interaction would be part of a **6**, the part that becomes a singlet when $SU(3)$ is reduced to $SO(3)$. I mention this scheme mainly to show to what lengths one has to go to associate the spin $\frac{1}{2}$ haptlons of $N = 8$ supergravity with quarks and leptons. In this theory, we probably have to admit that quarks and leptons are composite.

Such a proposal was made by Ellis, Gaillard, and Zumino in 1980. I was working on similar ideas in the next office, but I didn't believe all their conclusions. What they did was very ingenious. They put $e = 0$, so that $SO(8)$ is not gauged, thus allowing the full global symmetry of chiral $SU(8)$. Then they supposed, as Cremmer and Julia had conjectured, that the $SU(8)$ somehow gauges itself dynamically. Next they assumed that the charge and the $SU(3)$ of color are not inside the $SO(8)$, but half in and half out of it. That way they didn't have to make the assignments symmetrical. For example, for the eight $J_z = \frac{3}{2}$ gravitinos, they were able to take the charges to be $-\frac{1}{3}$, $-\frac{1}{3}$, $+\frac{1}{3}$, $+\frac{1}{3}$, and four 0's. Then the first five constitute the **5** of $SU(5)$, and the last three constitute the **3** of a sort of highly broken family $SU(3)$, which is supposed to account (because 3×3 antisymmetric is $\bar{\mathbf{3}}$) for the three families of fermions.

Then they went through some less convincing arguments in order actually to drag out $SU(5)$ with three families; I pass over those in silence. A very important idea, however, is the suggestion that all the spin 1, spin $\frac{1}{2}$ and spin 0 particles in quantum unified dynamics are fake. None of them is a haplon. Of the familiar particles, only the theoretical graviton is a haplon. Quantum unified dynamics (including, of course, the standard Yang-Mills dynamics) is a phenomenological renormalizable field theory, and all the parameters in it are lumped parameters that express the dependence on the cutoff somewhere up near the Planck mass. Thus an explanation is offered of the complexity of unified dynamics, with its numerous parameters and apparently arbitrary representations.

This last idea they attributed to Veltman; I don't know whether Veltman admits having invented it. There is also 't Hooft who has worked on it, I am told, and Norton and Cornwall at UCLA, and perhaps others. I think it is an intriguing notion that none of today's elementary particles is elementary except the graviton. We will hear, by the way, I think from Steve Adler,* about the opposite notion that the graviton is composite and Yang-Mills fields are elementary. Maybe they are made up of each other and we're back to the bootstrap! That is not really a joke, as we shall see.

To return to $N = 8$ supergravity, I've mentioned the various attempts to connect it to experiment. Now let us describe some of its generalizations. $N = 8$ supergravity can be derived by dimensional reduction from two distinct $N = 2$ supergravity theories in 10 dimensions called A and B. We note also that A, but not B, can be derived from $N = 1$ supergravity in 11 dimensions. It is very tempting, therefore, as suggested by Scherk and Schwarz about eight years ago, to take 6 or 7 extra dimensions seriously and give them nonzero size. These dimensions probably have to be small, although it would be very interesting to know what the limits are on such dimensions from experimental physics and cosmology. I would prefer not to be constrained by the argument that a size R around the Planck length is required. That argument stems from the fact that the reciprocals of effective coupling constants arise from κ (the square root of Newton's constant) times these R s, and that is a possible way, of course, of ultimately calculating coupling constants. If we try to identify these coupling constants with familiar ones that are no smaller than 10^{-2} , we must have R no larger than 10^2 times the Planck length. But perhaps those coupling constants are not the familiar ones and are very small.

* Editors' note: The paper read to the conference by Adler appears in this volume.

Many theorists have followed up the idea of extra dimensions within the framework of "Kaluza-Klein theory." The extra dimensions, taken seriously, tend to give internal symmetries, even broken internal symmetries—and they are really internal! The extra dimensions are wrapped up in a tiny structure, and for the first time the name "internal symmetry," which I always thought was silly, is justified. The isometries of the structure form a gauge group. One can use not only a linear representation of some symmetry group, like $SO(7)$ in 7 dimensions, or $SO(6)$ in 6 dimensions, but also (as I'm sure we'll hear at great length at this conference) a nonlinear quotient space representation—for example, $SU(3) \times SU(2) \times U(1)$ divided by $SU(2) \times U(1) \times U(1)$, which gives 7 dimensions; or $SU(3) \times SU(2)$ divided by $SU(2) \times U(1) \times U(1)$, which gives 6 dimensions. It may seem that $SU(3) \times SU(2)$ is too small, but remember, in 10 dimensions we have $N = 2$ supergravity, and so we have an additional $U(1)$ left over, and we can get $U(1) \times SU(3) \times SU(2)$ after all. (This remark is due to John Schwarz.)

Ideally the "compactification" of the extra dimensions will occur spontaneously, as a result of the equations of motion in the large space. Happily, mechanisms of spontaneous compactification are being found and will surely be discussed at this meeting.

Finally we go to string theories, which are particularly beautiful. You remember what a string theory is: it has an infinite number of states with increasing angular momenta lying on an infinite number of initially linear Regge trajectories. Early string theories, starting with the Veneziano model, were invented to describe hadrons according to the bootstrap idea. But then we developed QCD, and as far as hadrons were concerned, strings were seen as providing only a crude approximation to QCD. For hadronic purposes, of course, the slope of each Regge trajectory was around $1 (\text{GeV})^{-2}$.

The Veneziano model contains only bosons, and it also predicts a state of negative mass squared, a difficulty that has never been convincingly overcome by any explanation in terms of an unstable vacuum. But around 1970, Neveu and Schwarz found a string theory with bosons and fermions; important work on its properties was done at that time also by Ramond. The "critical dimension" for that string theory is 10, which means it is known to work in 10 dimensions, and might possibly exist in fewer dimensions if some conjectures of Polyakov and others turn out to be right; I shall assume here that it requires 10 dimensions.

A few years later, Scherk and Schwarz did further work on the Neveu-Schwarz string theory and adapted it to an entirely different task. They

made a slight modification in the slope of the Regge trajectory, by a factor of 10^{38} , and adapted it for use in gravitation. It was later shown that the theory didn't have any problems: that there was no negative mass squared and no negative probability, assuming a modification suggested by Gliozzi, Scherk, and Olive.

The Neveu-Schwarz theory is supersymmetric, and indeed their work and that of Ramond had anticipated in a sense the invention of supersymmetry. Their old superstring, when restricted to massless states in 4 dimensions, was now shown to give $N = 4$ supergravity with $N = 4$ super-Yang-Mills matter. But, treated in 10 dimensions, it was renormalizable—string renormalizable to at least one loop. I say string renormalizable because the string theory is so far treated only on the so-called mass shell. It hasn't yet got the regular Lagrangian formalism, although there's no reason to believe that there isn't one, and Green and Schwarz are working on it. The full Lagrangian formalism may be available soon. Anyway, although string renormalizable in 10 dimensions, the theory is hideously divergent, of course, when we restrict to 4 dimensions, because we have supergravity with supermatter. But this $N = 4$ super-Yang-Mills theory by itself is finite, as you will hear from West.

More recently, Brink, Green, and Schwarz have brought us "Superstrings II." This theory was not considered years ago because it doesn't have any open strings, and at that time people wanted open strings and closed strings for hadron purposes. Because Superstrings II has no open strings, it has no external supermatter when trivially reduced to initially massless states in 4 dimensions, and lo and behold, it becomes $N = 8$ supergravity! So this theory, which started from nothing but the idea of a string with bosons and fermions, fixes 10 dimensions, and then when you restrict it to zero mass and to 4 dimensions, fixes $N = 8$ supergravity as the limit. If you stop on the way, with the massless states in 10 dimensions, you get $N = 2$ supergravity. And just as there are two forms, A and B, of $N = 2$ supergravity in 10 dimensions, one of which can be obtained by reduction from 11 dimensions and one not, so there are two forms of Superstrings II—IIA and IIB. These theories are miraculously finite to one loop and may well be finite to every order, and that's rather exciting. Thus we have as advantages that the number of dimensions is fixed, that supergravity comes out and doesn't have to be put in, and that the theory may be finite to all orders. But we're left with the fundamental question of how superstrings are related to the real world.

Let me add two remarks, one on the cosmological constant and one on

the inflationary cosmology. You'll hear about both of these from lots of speakers, so I won't have to say much. I think the cosmological constant problem is one of the key issues of fundamental physics, and it's another question of renormalization. Renormalization, finiteness, and calculability are the whole story at our meeting here, as far as I can see, and that's entirely appropriate for a sequel to the 1947 Shelter Island Conference. Here we deal with renormalization of zero-point energy; that's something we always used to pass over very quickly in our books and classes: we put in some funny little points, and said it was gone. But GRAVITY IS WATCHING! It notices when we take away the zero-point energy, including the higher order corrections to that energy, and it says, "They've taken away a term in $\delta_{\mu\nu}$ from the $\Theta_{\mu\nu}$ tensor, and therefore they've added a term in $\delta_{\mu\nu}$ to Einstein's equation for me"—and that's a cosmological constant. The astronomical cosmological constant, which may or may not be identically the same quantity as the one in the equation (that's a problem of renormalization also, in a certain sense) is known to be zero or very close to zero. In natural units, it's 10^{-118} or smaller, natural units being those involving the Planck scale. That's the largest discrepancy in physics; it's large even for astrophysics.

(Gamow used to make fun of the big mistakes that astrophysicists tolerate. He wrote a letter about something or other in astrophysics and cosmology and deliberately made an error, preparing an erratum in advance. He sent the letter to the *Physical Review*, and then after it was printed he sent in the erratum, which said roughly, "In Equation so-and-so, there is a mistake by a factor of 10^{24} on the right-hand side. This does not affect the result.")

Anyway, 10^{118} bothers even astrophysicists. In supersymmetry theory (and this was originally one of the great things about supersymmetry) zero-point energy can be made to vanish to all orders if you include a constraint called R symmetry. However, when you put in spontaneous violation of supersymmetry, you get back a cosmological constant. Then, in supergravity theory, if you put in a self-coupling ϵ , so that you have gauging of $SO(N)$, you get another contribution with the opposite sign. Maybe they cancel—but nobody has yet found a way to cancel these terms naturally. If one simply dials the algebraic sum to zero, one is introducing the largest fudge factor ever.

Of course, if supersymmetry is explicitly violated, the algebra is altered. That's another possible source of the cosmological constant.

Then Hawking has developed a whole new way of looking at these matters, which I'm sure he will discuss,* in which the quantum fluctuations of gravity or supergravity, including the topology of the space, result in bubbles in space-time of Planck length size, forming a foam, which, he says, disguises a huge fundamental cosmological constant as a zero effective cosmological constant for the universe as a whole, thus divorcing the astronomical from the fundamental. Well, we'll see how well that works; it's an extremely ingenious idea.

Then we shall hear from Linde and from Guth about the remarkable new explosive or inflationary cosmology that accomplishes a gigantic increase in entropy of the universe during its early moments and dilutes out all sorts of unwanted things: monopoles, asymptotic curvature, inhomogeneity, and anisotropy, while apparently explaining the horizon paradox as well. All this by a mere addition of 10^{87} to the entropy. A very beautiful idea, in my opinion. The transition is accomplished specifically by a phase change in the vacuum, which is usually attributed to the quantum unified dynamics around 10^{15} GeV, although there might be other possibilities. Perhaps people will want to explore the idea that it occurs at an earlier era, or at least at a higher energy, and involves gravity more intimately, or supergravity. Possibly there are even extra dimensions—sizable extra dimensions comparable with the others—at the moment when this happens. Or superstrings: since these have huge entropies, they might furnish an alternative way to generate the big jump in entropy.

I understand that if the phase transition uses quantum unified dynamics, there is a slight discrepancy today with the usual $SU(5)$ theory, involving a difficulty with the size of the fluctuations producing galaxies. But that's highly technical and might be reparable by some variation in the theory.

Let me raise one question that may be answered at this meeting, and that is whether there is some inconsistency between Hawking's idea of divorcing the cosmological constant in the equation from the one in astronomy and the idea in the exploding universe theory that the cosmological constant jumps from a finite value to zero. (One doesn't explain the zero, only the jump.) For the exploding universe it's both the fundamental constant and the astronomical one that jump, and perhaps we can understand that better

* Editors' note: The paper read to the conference by Hawking appears in this volume, as do the papers by Linde and Guth mentioned in the following paragraph.

after a few days. Most likely there is actually no problem, but I don't understand it very well.*

Let me say in conclusion that I'm delighted that, during the last decade or more, an old prediction of mine has been confirmed; that particle physics and cosmology would essentially merge into one field, the field of fundamental physics, which underlies all of natural science. In our attempts to understand the basic structure of the universe, we theorists of fundamental physics, even though our day-to-day labors are often frustrating and petty like anyone else's, are engaged in a magnificent quest, along with our experimental friends, for a kind of Holy Grail; a universal theory. Will it prove as elusive as the Holy Grail? Will there be a Lancelot who almost grasps it? A Galahad to whom it is fully revealed? Whether or not we achieve the quest, each time we slay a dragon or rescue a maiden along the way, each splendid adventure is an accomplishment in itself; like the writing of a poem or a symphony, part of the soaring of the human spirit.

* Editors' note: Hawking makes the following interjection: "There is no problem!" Gell-Mann: "That's what you told me a few weeks ago; I'll be delighted to hear it explained again. Anyway, some day we may have a consistent picture of a very early universe, somehow trading its non-zero cosmological constant for zero to an accuracy of 10^{-12} ", and producing a lot of matter, including us."