UNIFICATION OF COUPLINGS

Recent high-precision experimental results support the predictions of the minimal supersymmetric SU(5) model that unifies electromagnetism and the weak and strong interactions.

Savas Dimopoulos, Stuart A. Raby and Frank Wilczek

Ambitious attempts to obtain a unified description of all the interactions of nature have so far been more notable for their ingenuity, beauty and *chutzpah* than for any help they have afforded toward understanding concrete facts about the physical world. In this article we wish to describe one shining exception: how ideas about the unification of the strong, weak and electromagnetic interactions lead to concrete, quantitative predictions about the relative strengths of these interactions.

The basic ideas in this subject are not new; they were all essentially in place ten years ago. For several reasons, however, the time seems right to call them back to mind. Most importantly, the accuracy with which the relevant parameters have been determined experimentally has improved markedly in the last few months, making a much more meaningful comparison between theory and observation possible. The results of this confrontation, as we shall see, are quite encouraging and suggestive.

Gauge theories

It has been traditional to identify four fundamental interactions: strong, weak, electromagnetic and gravitational. In the 1960s and 1970s great progress was made in elucidating the principles underlying the first three of these interactions (see the articles by Howard Georgi and Sheldon Lee Glashow in PHYSICS TODAY, September 1980, page 30, and by David Gross, January 1987, page 39).[1] By comparison the elucidation of quantum gravity is at a comparatively primitive stage. Except for a few remarks toward the end, our discussion will be confined to the first three interactions—the traditional domain of high-energy physics.

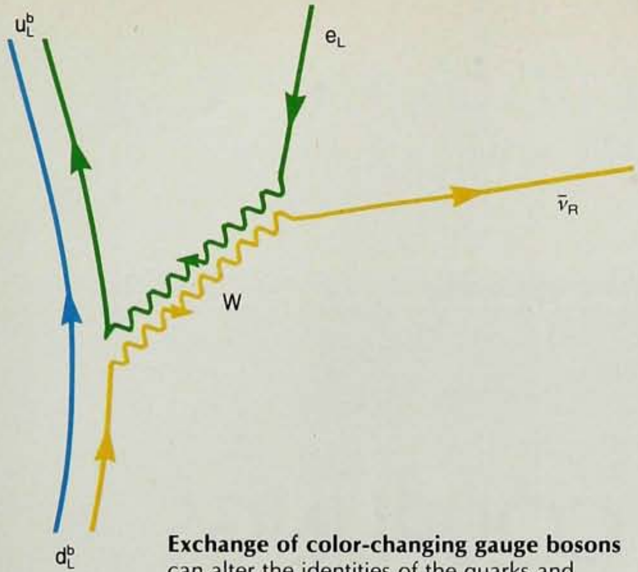
Savas Dimopoulos is a professor of physics at Stanford University, in Stanford, California. **Stuart Raby** is a professor of physics at Ohio State University in Columbus, Ohio. **Frank Wilczek** is a professor in natural science at the Institute for Advanced Study, in Princeton, New Jersey.

To make a very long story short, it was found that a common mechanism underlies all three of these interactions: Each is mediated by the exchange of spin-1 particles, *gauge bosons*. The gauge bosons have different names in the three cases. They are called color gluons in the strong interaction, photons in the electromagnetic interaction, and W and Z bosons in the weak interaction. But despite the difference in names and some other superficial differences, all gauge bosons share a common mathematical description and deeply similar physical behaviors. Gauge bosons interact with quarks and leptons in several ways—mediating forces among them, being emitted as radiation when the quarks or leptons accelerate, and even changing one kind of quark or lepton into another.

The original and most familiar gauge theory is also the most basic. Quantum electrodynamics is properly understood, in modern terms, to be neither more nor less than the theory of a single gauge boson (namely, the photon) coupled to a single charge, or "color" (namely, electric charge). In mathematical language, it is the theory of the gauge group U(1).

Chromatic terminology for charges is useful and evocative, but must not be taken too literally. Color charges are numerical quantities, which may be positive or negative integers (or zero). The charges associated with different colors are independent quantities. Thus a particle may carry blue charge $+1$ and yellow charge $+1$ but green charge 0.

The modern theory of the weak interaction is essentially the simplest nontrivial extension of this setup, to include *two* colors. An important new possibility for gauge boson physics first shows up with two colors: In addition to gauge bosons that, like the photon, respond to the color charges, there are also gauge bosons that change a unit of one charge into a unit of the other. In this fundamental process (see figure 1), one kind of particle is changed into another carrying different color charge. Color charge is



Exchange of color-changing gauge bosons can alter the identities of the quarks and leptons involved. Here, the elementary process underlying ordinary radioactivity is depicted as a process of weak color transmutation. **Figure 1**

conserved overall because the difference in charge between the altered particles is carried by the gauge boson. The W bosons are of this identity-altering type, and their exchange is the mechanism underlying radioactive transmutations of atomic nuclei of one element into those of another. The Z boson, acting more like the photon, responds to but does not alter the weak color charges. In mathematical language, the modern theory of the weak interaction is the theory of the gauge group SU(2)—the 2 here just indicates two colors.

Finally quantum chromodynamics, the modern theory of the strong interaction, is—you guessed it—the theory of three colors, based on the gauge group SU(3). It involves eight gauge bosons (color gluons), six that alter colors and two others that merely respond to them.

The color charges involved in the strong and weak interactions are completely distinct. It has become customary, at least in the US, to call the strong colors red, white and blue. The weak interaction gives us an opportunity to soften the chauvinism of this terminology to some extent, by adding two new colors: Call them yellow and green. It might seem that to complete the structure we would need a sixth color, for electromagnetic charge. But most remarkably, it appears that having identified the five strong and weak colors, we do *not* need to add a sixth, separate color for electromagnetism. Electric charge is not independent of the other charges. If we make the color assignments indicated in figure 2 (whose true significance will emerge only below), then the electric charge Q of a particle is given in terms of its various color charges (R, W, B, Y and G) according to the simple formula

$$Q = -\tfrac{1}{3}\,(R + W + B) + G \qquad (1)$$

Unification: Triumphs and challenges

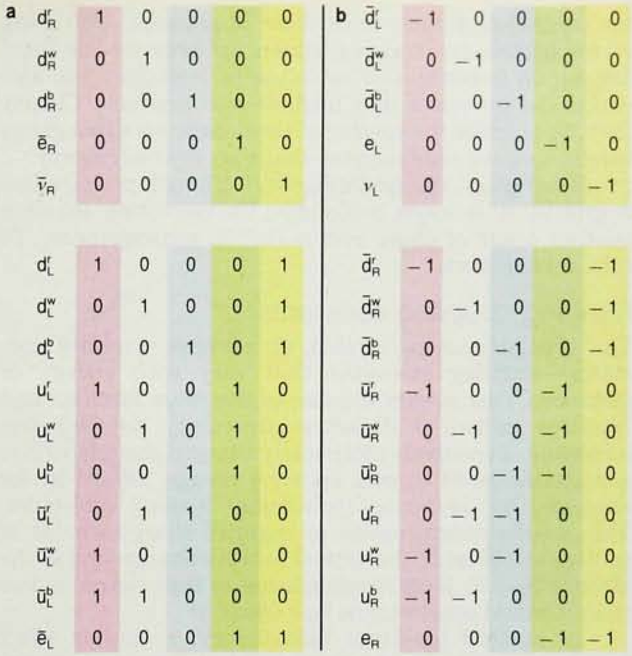
The fact that all three major interactions of particle physics can be described using the concept of gauge bosons coupled to color charges hints at some deeper unity among them. So too, with more subtlety and power, does equation 1. The strong color gluons mediate all possible changes and responses among the red, white and blue colors, while the weak gauge bosons do the same between the yellow and green colors. What could be more natural than to postulate the existence of gauge bosons corresponding to all possible changes and responses among *all five* colors?[2] Such bosons would include the color gluons, weak bosons and photon, and also additional gauge bosons that would change (for example) red charge into yellow charge. Altogether 12 new gauge bosons must be added to the 12 known ones. The gauge theory for five colors is denoted SU(5). It includes the $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$ gauge theories of the strong, weak and electromagnetic interactions—and more.

This idea, on cursory examination, suggests two lovely qualitative successes and two quantitative disasters.

First, the successes. If we consider only the gauge bosons, the expansion of the theory appears as an appealing but quite speculative possibility. While it suggests the existence of new gauge bosons, it does not shed much light concerning the properties of the ones we already know to exist. However, if we widen our considerations to include quarks and leptons, a wonderful advantage of the larger theory comes into view. As indicated in figure 2, the 15 quarks and leptons within a family can be grouped into two classes. One class consists of five particles, each carrying one unit of one of the five color charges. The other class consists of ten particles, each carrying a unit of each of two distinct color charges. Within either of these two classes, transformations carrying any given particle into any other one can be mediated by appropriate gauge bosons. In other words, the particles within either class are all related to one another by the gauge interaction. They are like different faces of a single die—inseparable, symmetrical pieces of a larger whole.

In mathematical terms, the particles fall into two irreducible representations of SU(5): a five-dimensional vector representation and a ten-dimensional antisymmetric tensor representation. By contrast, when we restrict ourselves to the transformations of $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$, the particles in a family fall apart into no less than five different classes. This striking gain in economy of description is one great qualitative success of the simplest SU(5) unification scheme.

The other success concerns equation 1. This marvelous equation, in which the electromagnetic, strong and weak charges all come into play, was an encouraging hint toward unification. Within SU(5), its potential is brilliantly fulfilled. Although it is a little too complicated for us to derive here, it is not terribly difficult to show that equation 1 is an automatic consequence of unification in



a

d^r_R	1	0	0	0	0
d^w_R	0	1	0	0	0
d^b_R	0	0	1	0	0
$\bar{e}_R$	0	0	0	1	0
$\bar{\nu}_R$	0	0	0	0	1
d^r_L	1	0	0	0	1
d^w_L	0	1	0	0	1
d^b_L	0	0	1	0	1
u^r_L	1	0	0	1	0
u^w_L	0	1	0	1	0
u^b_L	0	0	1	1	0
$\bar{u}^r_L$	0	1	1	0	0
$\bar{u}^w_L$	1	0	1	0	0
$\bar{u}^b_L$	1	1	0	0	0
$\bar{e}_L$	0	0	0	1	1

b

$\bar{d}^r_L$	−1	0	0	0	0
$\bar{d}^w_L$	0	−1	0	0	0
$\bar{d}^b_L$	0	0	−1	0	0
e_L	0	0	0	−1	0
ν_L	0	0	0	0	−1
$\bar{d}^r_R$	−1	0	0	0	−1
$\bar{d}^w_R$	0	−1	0	0	−1
$\bar{d}^b_R$	0	0	−1	0	−1
$\bar{u}^r_R$	−1	0	0	−1	0
$\bar{u}^w_R$	0	−1	0	−1	0
$\bar{u}^b_R$	0	0	−1	−1	0
u^r_R	0	−1	−1	0	0
u^w_R	−1	0	−1	0	0
u^b_R	−1	−1	0	0	0
e_R	0	0	0	−1	−1

Quarks and leptons collect into two classes when assigned strong and weak colors inspired by SU(5) unification (**a**). We label the strong colors red, white and blue and the weak colors green and yellow. Reading across each row gives the SU(5) color charge for each particle. Subscripts L (left) and R (right) indicate the chirality of the particles, while the superscripts r, w and b indicate the standard strong-color labels of the u (up) and d (down) quarks. Overbars indicate antiparticles (**b** explicitly displays the antiparticles of the particles in **a**). Notice that left-handed and right-handed versions of the same particle may be differently colored—this reflects the violation of parity in the weak interactions. **Figure 2**

SU(5). The photon only fits within this symmetry group if it responds to precisely the combination of color charges that occurs in equation 1. Thus unification offers a framework in which the apparently chaotic spectrum of electric charges of quarks and leptons can be understood rationally.

In a more precise treatment we would actually have to worry about the spectrum of weak hypercharges, which is even worse. One of us (Wilczek) recalls that as a graduate student he considered the now standard SU(2)×U(1) model of electroweak interactions to be "obviously wrong" just because it requires such ugly hypercharge assignments. That was going too far, but it still seems fair to call the model "obviously incomplete" for this reason.

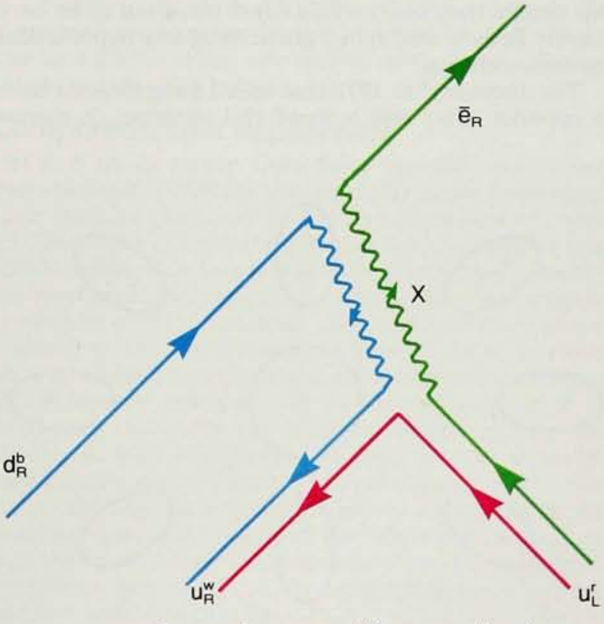
Now we must describe two daunting difficulties that such attempts at unification face. The first disaster is that the different gauge bosons, although they do similar things, do not do them with the same vigor. In other words, they couple to their respective color charges with different strengths. The strong interaction, as befits its name, really is much stronger than the weak interaction, which in turn is slightly stronger than the electromagnetic. Thus the perfect symmetry among colors required in a truly unified gauge theory doesn't seem to be in the cards. We shall return to this problem below.

The second disaster concerns the processes mediated by the extra gauge bosons, particularly the ones that change strong into weak color charges. It is always at least slightly worrisome to postulate the existence of hitherto unobserved particles, but these fellows are especially objectionable, because their exchange mediates, through the mechanism indicated in figure 3, processes capable of destabilizing protons. However, protons are rather reliably reported not to decay. Even in 1974, when unified theories of the type we are discussing were first proposed, the lifetime of the proton was known to be upwards of 10^{21} years. Since then, systematic experiments have raised the lower limit to over 10^{31} years (for most plausible decay modes).[3] Comparing this to the rates for comparable weak decays, which are measured in microseconds, we realize with a start what an enormity is being perpetrated—these new gauge bosons must be indeed very different from, and in some sense much less potent than, the old (that is, known) ones.

Nevertheless, given the qualitative successes of gauge theory unification, and its ineluctable beauty, one must not give it up without a fight. And indeed both difficulties can be overcome in triumphal style.

Let us take the second difficulty first. It is actually not so difficult to explain this problem away. To do so, we must now mention a very important aspect of gauge theories that for simplicity we have so far neglected: These theories may exist in different phases and exhibit properties at low energies that differ somewhat from their symmetrical high-energy behavior. For our purposes, the most important point is that gauge bosons may become massive, through the so-called Higgs mechanism,[4] and the heavier the gauge bosons, the rarer are the processes mediated by their exchange. (The Higgs mechanism is in a very direct sense simply a relativistic version of Fritz and Heinz London's superconducting electrodynamics.)

This, by the way, is why the weak interactions are much less prominent than electromagnetism, even though the intrinsic strengths of the weak-vector-boson couplings



Proton decay would be caused by the exchange of some of the extra gauge bosons (X) needed for unification. These bosons would change strong into weak colors and would lead to processes wherein three quarks change into an antilepton. **Figure 3**

are somewhat greater than those of photons. The weak vector bosons are massive, which not only makes them difficult to produce and unstable in isolation, but also makes the processes they mediate less vigorous. Clearly then, to exorcise the specter of the dangerous extra gauge bosons we need only suppose that they are very heavy.

What about the first difficulty? Though perhaps less dramatic, it is more profound. Its resolution involves another order of ideas, and is rich in consequences. To this, we now turn.

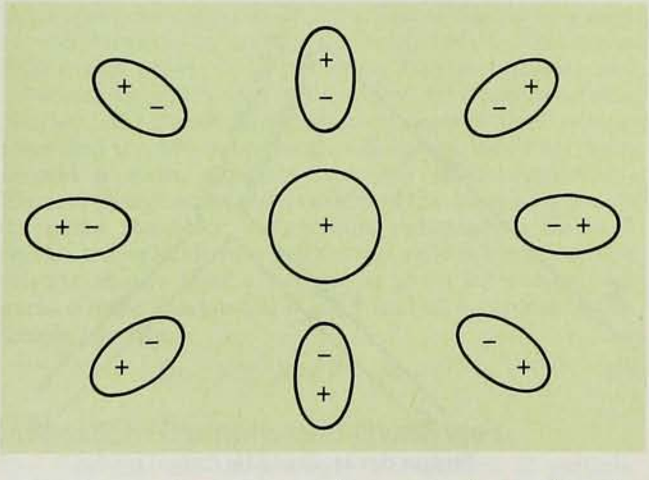
Running coupling constants

The crucial concept is that of running coupling constants—coupling strengths that vary with energy or distance. This is very similar to the more familiar and intuitive notion of dielectric screening. In dielectric screening, a positively electrically charged particle within a material tends to pull negative charge toward it, for example, by distorting (polarizing) neutral molecules. This nearby enhancement of negative charge shields or screens the effect of the central positive charge, and so the electric field at large distances due to that charge is less than it would otherwise be (see figure 4).

In modern quantum field theory, a similar effect happens even in empty space. This is because "empty space" is not a true void, but rather a dynamical medium full of virtual particle–antiparticle pairs that flicker briefly into existence and then reannihilate before traveling very far. (A less poetic but still visually appealing view of the same thing is afforded by the Feynman diagrams in figure 5.) These denizens of the vacuum can be polarized, no less than molecules in a solid. As a result the charge and electric field distributions close to a nominal elementary "point particle" are in fact structured: The charge is partially screened. The vacuum is a dielectric.

Ordinary dielectric screening tends to make the effective charge smaller at large distances. Conversely, of course, if we work from the outside in, we see the effective charge gradually increasing from what we saw from far away. Virtual quarks and leptons also tend to screen any color charge they carry. This effect turns out to be very general: Spin-$\frac{1}{2}$ and spin-0 particles of any hypothetical type screen charge.

The discovery[5] in 1973 that spin-1 gauge bosons have the opposite effect was a wonderful surprise. It means that a charge that looks large and formidable at large distances can be traced to a weaker (and calculationally more manageable) source at short distances. The discovery of this dynamical effect—known as asymptotic freedom—led directly to the identification of SU(3) color gauge theory, or QCD, as the theory of the strong interaction.[6] This came about because the SLAC electroproduction experiments,[7] demonstrating the phenomenon of scaling, indicated that the strong interaction between quarks is much weaker at short distances than one would infer from afar. More precisely, what these experiments indicated is that rapidly accelerated quarks emit few gluons. In other words they behave when they are hit hard as if they are ideal structureless point particles: They recoil elastically; they have no "give." This behavior is in contrast to their appearance when hit softly. Then the more powerful, longer-range aspect of the strong interaction causes quarks to behave not like points but more like thick balls of virtual gluons, quarks and antiquarks.



Dielectric screening occurs when a charge in a dielectric medium polarizes the molecules around it. This cloud of polarization partially hides, or screens, the central charge. **Figure 4**

The logic of the discovery of hard structureless particles—Richard Feynman's partons,[8] now identified with quarks and gluons—inside the proton in the SLAC experiments is quite similar to the logic underlying the classic experiment of Johannes Geiger and Ernest Marsden. Their observation that alpha particles impinging on gold foil may be violently deflected through large angles was interpreted by Ernest Rutherford as indicating the existence of hard, effectively point-like nuclei at the center of atoms. Replacing alpha particles with electrons, and nuclei with partons, we essentially map the Geiger–Marsden experiment onto the SLAC experiment.

Later experiments, as we shall discuss further below, have confirmed and sharpened the early indications from SLAC. When quarks are rapidly accelerated they usually propagate exactly as ideal structureless point particles, but occasionally radiate one or more color gluons instead. QCD gives a detailed quantitative account of these matters, and has been very successful in predicting the outcomes of experiments[9] (so much so, that experimentalists now rely on it to calculate their backgrounds).

Why do these bosons have the opposite effect from other particles? The mechanism of screening seems so clear and inevitable that its reverse seems implausible. However, it turns out, roughly speaking, that attractive magnetic dipole–dipole attractive forces between like-charge gauge gluons outweigh their electric repulsion, leading to an accumulation of the *same* charge—antiscreening![10]

In our present context, it is convenient to consider screening and asymptotic freedom as functions of energy rather than distance. In a sense that can be made precise, in quantum mechanics high energy or momentum corresponds to small distance. Roughly speaking, then, screening corresponds to the coupling's increasing with energy, while asymptotic freedom corresponds to its decreasing.

The coupling of SU(3) is more affected by asymptotic freedom than are the other couplings, simply because there are more strong color gauge bosons. It outweighs the effect of the quarks. For weak SU(2) the competition is more equal, while for electromagnetic U(1) there is no gauge boson contribution, and ordinary screening wins. As a result the strong coupling decreases at large energies, while the weak stays nearly constant and the electromagnetic increases. But these are just the directions of change that can cause the couplings to merge!

This whole circle of ideas is beautifully summarized in the plot of effective couplings against energy or mass scale due to Howard Georgi, Helen Quinn and Steven Weinberg[11] (figure 6). The energy scale for the running of the couplings is logarithmic, so it takes a big change in energy to see any change in the couplings. Thus the scale at which unification takes place will be very much larger than what we are accustomed to in accelerator physics.

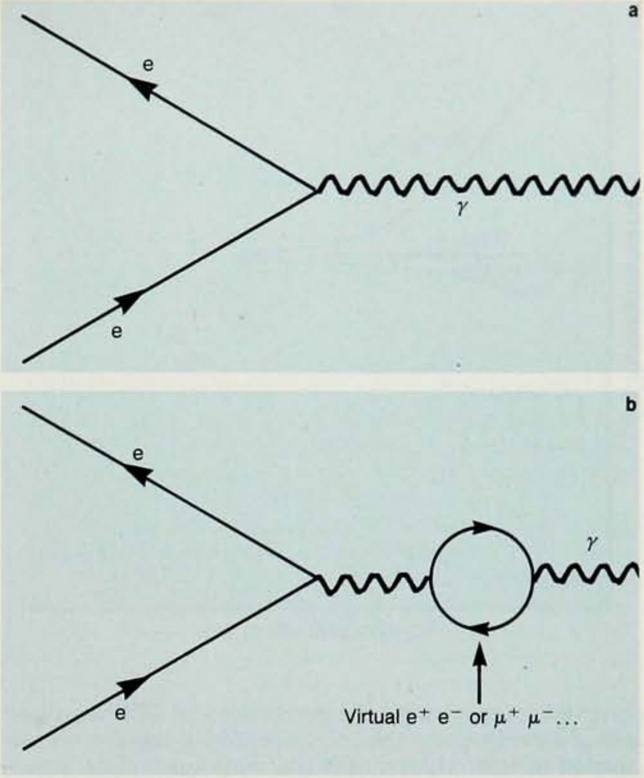
The logic of prediction from unification

A method for comparing unified theories and reality, using the observed strength of couplings, emerges from careful consideration of figure 6. On the left-hand side of the plot, we have three measurable parameters: the strong, weak and electromagnetic couplings. On the right-hand side, we have two unknown parameters: the mass scale for restoration of the full unified symmetry, and the strength of *the* coupling (there's just one!) when this occurs. Since the three measurable parameters are supposed to derive from two more fundamental (but *a priori* unknown) ones, they must obey a constraint. *The primary prediction from the logic of unification is a numerical relationship among the strong, weak and electromagnetic couplings.*

The form of this relationship is conveniently expressed in terms of the weak angle θ_W defined by the expression $\tan\theta_W = g_1/g_2$, where g_1 and g_2 are the coupling strengths of the U(1) and SU(2) gauge bosons. The coupling g_2 is directly the coupling strength of W bosons, analogous to the electromagnetic coupling e for photons. The physical photon is a mixture of the fundamental U(1) and SU(2) gauge bosons, and one finds $e = g_2 \sin\theta_W = g_1 g_2/(g_1^2 + g_2^2)^{1/2}$. Given the experimental values of the strong coupling g_3 and the electromagnetic coupling e, the logic of unification allows one to predict the value of $\sin^2\theta_W$, which is the quantity experimentalists generally report. The precise value predicted depends on the spectrum of virtual particles that enters into the calculation of the running of the couplings. We shall elaborate on the numerical aspect of these predictions and their comparison with experiment in a moment.

The predicted constraint on the observed couplings, however, does not exhaust the interest of this calculation. If the observed couplings do obey the constraint, we will also obtain definite predictions for the mass scale of restoration of symmetry and for the value of the coupling at this scale. Together these allow us to predict the mass of the dangerous gauge bosons whose exchange destabilizes protons and to obtain a rate for proton decay through this mechanism, which in principle (and perhaps in practice) can be compared with experiment. In the context of cosmology, these parameters determine the temperature at which a phase transition from unbroken to broken unified gauge symmetry occurred during the Big Bang.

Slightly more subtle but perhaps in the long run even more important for the future of physics is another aspect of the unified coupling and scale. A classic problem of physics for the past several decades has been the meaning of the numerical value of the fine-structure constant $\alpha = e^2/4\pi\hbar c$. This pure number largely controls the structure of the world (that is, all of chemistry and most of physics, as Dirac described the domain of quantum electrodynamics). Many attempts, ranging from crackpot numerology to serious efforts by leading physicists, have been made to calculate its value from deeper principles. None has succeeded. Unified theories radically alter the terms of this problem but do not remove its substance. It becomes, if anything, grander. The fine-structure constant no longer appears as a simple primary ingredient of fundamental theory. Rather it, together with the strong and weak couplings, derives from the primary unified coupling at short distances by processes of renormalization and symmetry breaking. The right problem, it seems, is not to try to calculate α but rather to explain why the unification coupling and scale are what they are.



Empty space is a dielectric medium in quantum field theory and can screen charge. These Feynman diagrams represent the interaction between a bare charge and a photon (**a**) and the effect of virtual particles coming between the photon and the charge (**b**). **Figure 5**

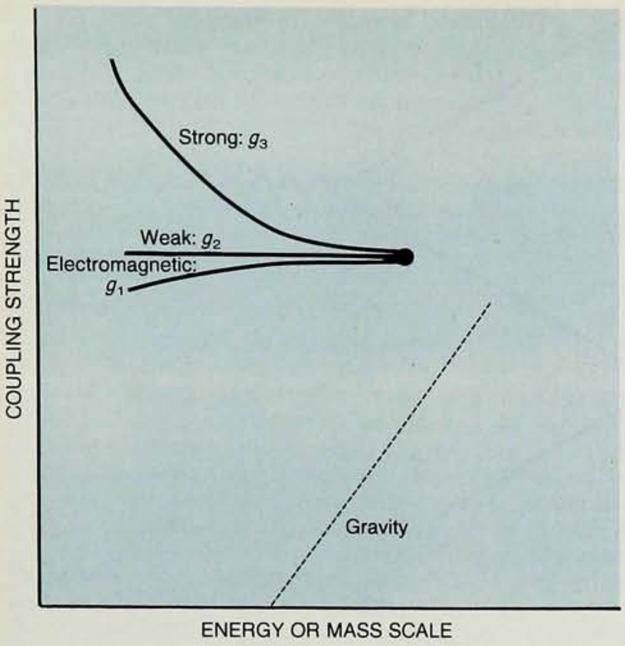
Finally, the value of the scale of unification has important implications for the eventual reconciliation of gravity with the other, now unified interactions. We shall return to this point below.

Comparison with experiment

But first let us return from these rarefied heights back down to earth, to discuss experimental measurements and their confrontation with models in a more concrete way.

The value of the strong interaction coupling (at some definite scale) can be measured in many ways. Perhaps the most intuitively appealing is to use electron–positron annihilation into hadrons. The fundamental process underlying the annihilation is production of a virtual photon, which converts to a quark–antiquark pair (figure 7). Of course one does not see actual quarks in the laboratory, but only the hadron showers or jets they induce. At high energies the dressing process, whereby a bare quark is converted into physical hadrons, is soft. This means that the hadrons in the jet are all moving in very nearly the same direction as the underlying quark, and that the total energy and momenta of the particles in the jet add up very nearly to the underlying quark's energy and momentum (figure 7a).

However, there is also a small probability—proportional to the strong coupling—for the quarks to radiate a hard gluon, that is, one with substantial energy and momentum of its own. In that case one should observe events with three jets, as shown in figure 7b. Such events are indeed observed. That their angular distribution is



The running of the strong, weak and electromagnetic couplings, extrapolated to very high mass scales, can result in their meeting at a point. This occurs if these three interactions, as observed at low energy, all result from the spontaneous breakdown of a unified theory at a large mass scale. The gravitational coupling, shown schematically on this plot, starts out very much smaller than the other couplings but would join them if extrapolated to about 10^{19} GeV. **Figure 6**

observed to agree with the predictions of QCD provides splendid evidence for the existence of spin-1 gauge bosons coupled to color charge with the same space–time structure as the coupling of the photon to ordinary charge. Most importantly for our purposes, the ratio of three-jet to two-jet events gives a direct quantitative measure of the strength of the strong coupling. Four-jet events are also observed at the expected (low) rate.

Conceptually, the most straightforward way to measure the weak coupling is simply to measure the mass of the W boson. Indeed, the rate of all weak processes at low energy, including, for example, the easily-measured rate of muon decay, is governed by the ratio of the weak coupling to the mass of the mediating W boson. (This follows from graphs like that shown in figure 1, using the elementary Feynman rules.) In practice other, more complicated measurements, involving the mixing of the Z boson with the photon, are more easily made accurate.

The electromagnetic coupling, of course, has been known with extreme accuracy for a long time.

Though no true ambiguities arise if one uses the theory to calculate physically meaningful quantities, quantitative comparison with experiment requires great care. For instance it is not at all trivial to define the couplings properly and consistently, because they all run and in any physical process there are a variety of ways one might choose the nominal "scale." Once the experimental measurements are properly translated into values of the couplings, it becomes possible to confront them with the predictions of different unified models.

This involves at least two criteria. First, we must demand that the constraint on the observed couplings is satisfied. Second, we must demand that the predicted rate of proton decay through gauge boson exchange is not too large. Different models of unification will contain different numbers and kinds of virtual particles, which will cause the couplings to run differently. Therefore, in general, different models will lead to different constraints on the observed couplings and to different values for the unification scale and coupling.

It is good scientific strategy to check the simplest possibility first. The simplest unified model is the one based on SU(5) unification as described above. The minimal version of this model does not require any unobserved particles with mass significantly less than the unification scale beyond those already needed in $SU(3)\times SU(2)\times U(1)$ without unification (that is, the top quark and the Higgs boson). This minimal model does amazingly well by the first criterion. Until quite recently the measured values of the couplings did satisfy the constraint imposed by this minimal unified model, within experimental uncertainties. This is a truly remarkable result and greatly encourages us to think that there is much truth captured within this circle of ideas. The minimal SU(5) unified model has difficulties, however, meeting the second criterion. It predicts too large a rate of proton decay or, equivalently, too small a unification scale. The predicted scale is roughly 10^{15} GeV, and the predicted lifetime is roughly 10^{29} years. This is not quite acceptable, as we mentioned before. On the other hand the predicted proton lifetime is not absurdly short—certainly it is a vast improvement on a microsecond!—and in fact special, heroic experimental efforts were required to rule it out.

Given these results, it seems wise merely to tinker with the basic ideas rather than simply junk them. Are there compelling alternatives to the minimal unified model? Do they manage to retain its successes while remedying its shortcomings?

Supersymmetry

Once we wander from the straight and narrow path of minimalism, infinitely many silly ways to go wrong lie open before us. In the absence of some additional idea, just adding unobserved particles at random to change the running of the couplings is almost sure to follow one of these. However, there are a few ideas that do motivate definite extensions of the minimal model and are sufficiently interesting that even their failure would be worth knowing about.

Surely supersymmetry[12] is in this class. In a well-defined sense, supersymmetry is the only possible way to unify the description of particles with different spins. Indeed, it is a symmetry whose basic operation is to transform particles or fields with one spin into other particles or fields whose spins differ by the minimal unit $\hbar/2$. In the process it transforms bosons into fermions, and vice versa. As yet there is no direct sign of supersymmetry in nature (the developments reported below are probably the nearest thing so far), and if supersymmetry is relevant to the description of nature it must be broken. However, a broken symmetry can still be rich in consequences if its breakdown occurs in a mild and orderly way.

Perhaps the most appealing idea in this direction is that the breakdown of supersymmetry is spontaneous.[13] This means that it remains a valid symmetry of the underlying *laws* of physics but is broken in the course of the evolution of the state of the universe. This process is similar to the way the alignment of spins in a ferromagnet spontaneously breaks rotational symmetry as the magnet is cooled through its Curie point: Rotational symmetry is still valid in a fundamental sense, even in a magnet, but the stable configurations of spins within the magnet do not respect it. The fundamental laws have more symmetry than any of their stable solutions.

Supersymmetry is a necessary ingredient in several other theoretical ideas. There are many hints that it may help to elucidate the gauge hierarchy problem—that is, the vast difference between the unification scale and the scale of electroweak symmetry breaking.[14,15] When one promotes supersymmetry to the status of a local gauge symmetry, one finds that Einstein's general relativity is a necessary consequence, thus finally bringing that theory within the circle of ideas used to describe the other interactions of particles.[16] And, of course, supersymmetry is a necessary ingredient in superstring theory, the most promising concrete approach to unifying the three interactions of particle physics with gravity currently known.

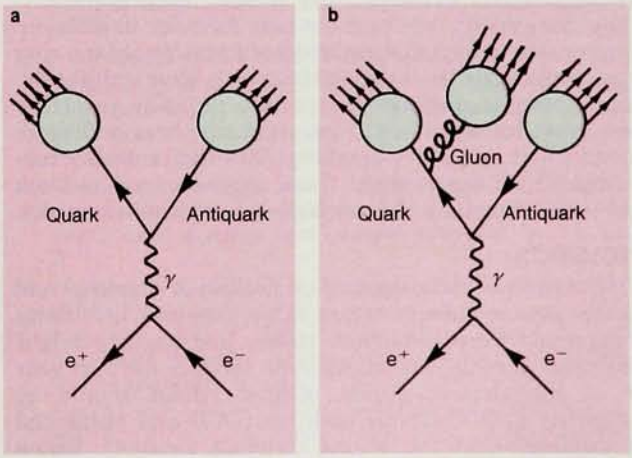
Pioneering attempts[17] to incorporate supersymmetry into realistic models of particle physics ran into various difficulties. Finally a consistent phenomenological model was found, using the idea of soft symmetry breaking.[18]

Ten years ago, we pointed out that extension of the minimal model to incorporate supersymmetry had important implications in the context of the ideas discussed above.[15] The most important change suggested by supersymmetry is that one should include many additional particles with masses of 10^4 GeV or less, whose properties are predicted with sufficient definiteness to allow a meaningful analysis of their effect on the running of couplings. Roughly speaking, one should expect a doubling of the spectrum of elementary particles at these energies. This doubling occurs because supersymmetry transforms particles into their superpartners, differing in spin by $\hbar/2$ but with closely related couplings, and none of the known quarks, leptons or gauge bosons can be identified with the superpartner of any other. The mass estimate for the superpartners is not quite firm, but if supersymmetry is to help address the hierarchy problem it seems necessary that its breaking (of which the mass difference between superpartners is a measure) not be too large.

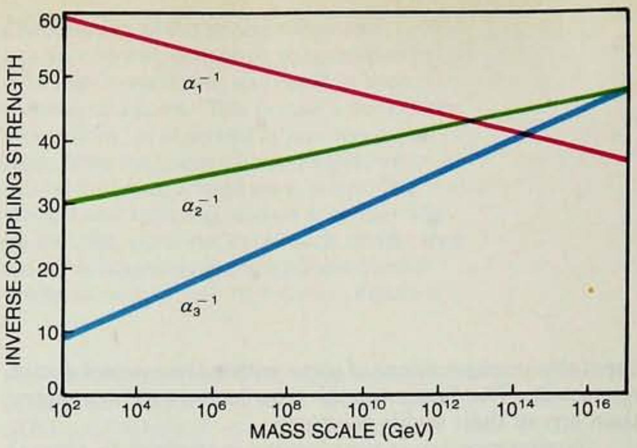
What are the effects of adding these superpartners? The main effect is to raise the scale of unification *without much disturbing the successful SU(5) relation among couplings.* Indeed, the main reason superpartners tend to raise the scale of unification is that the gluinos, the spin-$\frac{1}{2}$ partners of the color gluons, partially cancel the asymptotic-freedom effect of the gluons themselves. Thus it takes a longer run in energy for the biggest difference between couplings—the anomalously large strength of the strong interaction—to get wiped away. On the other hand the group theoretic structure of the calculation, which controls the ratio of couplings, is not much affected by the new superpartners. This is because the new superpartners occur in the same symmetrical pattern as their known counterparts. (Indeed supersymmetry relates particles with different spins but the same gauge—that is, color—charges.) Thus, roughly speaking, the running of each coupling is slowed down by the same factor.

Of course, by raising the scale of unification, the supersymmetric unified models made it seem less certain that the proton would decay on schedule. As time went on and no decays were observed, it became clear that this might be just what the doctor ordered.

Until recently it was appropriate to emphasize that incorporating supersymmetry into simple unified models does not drastically change the relation that they predict among coupling constants, since this prediction was consistent with the available data. However, small deviations from the nonsupersymmetric predictions do exist, because it is not quite true that particles (and their superpartners) occur in a completely symmetrical pattern.[15,18,19] The "bad actors" are the scalar fields introduced to implement electroweak symmetry breaking—the Higgs fields. In constructing the standard electroweak $SU(2)\times U(1)$ theory one must introduce a complex doublet of Higgs fields carrying the weak color charges. However, on phenomenological grounds one must not introduce their counterparts carrying strong color charge. Indeed exchange of the strongly colored Higgs particles can destabilize protons, and it leads to catastrophic rates for proton decay unless the mass of these particles is extremely large. There is no compelling understanding of why the strong-color Higgs particles are so heavy compared to their weak-color counterparts; this is one aspect of the gauge hierarchy problem. (Actually what is puzzling is not so much the heaviness of the strong-color Higgs particles but the lightness of the ordinary ones.) At



Electron–positron collisions at high energy can produce an elementary quark–antiquark pair, which materializes as two jets of particles moving in opposite directions (**a**). More rarely, the rapidly accelerated quarks radiate a color gluon, which produces a third jet (**b**). **Figure 7**



The most recent measurements of the low-energy couplings $(\alpha_i = g_i^2/4\pi\hbar c)$ clearly fail to meet at a point when they are extrapolated to high energies by computations incorporating the particle content of the minimal nonsupersymmetric SU(5) [or simply $SU(3)\times SU(2)\times U(1)$] model. This indicates that the observed low-energy couplings are not consistent with a unified model having just this particle content. The thickness of the lines indicates the experimental uncertainties. (Adapted from a figure provided by Ugo Amaldi, CERN.) **Figure 8**

present it seems wise to be pragmatic and simply accept nature's unequivocal indication that this is so. How does this mass difference affect the unification of couplings?

The normal, weak-color Higgs fields influence (to first order) only the running of the weak and, to a lesser extent, the electromagnetic couplings. They tend to make these couplings increase with energy. The inclusion of fermionic superpartners accentuates these effects. Furthermore, for technical reasons it turns out that in the minimal supersymmetric model one must introduce not just one but two weak Higgs doublets. The contribution of all the Higgs fields and their superpartners to the running of couplings is quantitatively small compared with the contribution of quarks, leptons and gauge bosons, but recent accurate measurements (especially the beautiful results from LEP[20]) can resolve the small corrections the Higgs fields are predicted to make for the constraint on the couplings. Ugo Amaldi, Wim de Boer and Hermann Fürstenau[21] conclude that the minimal supersymmetric model gives an excellent fit to the data, whereas the minimal nonsupersymmetric model is definitely excluded by many standard deviations. Figures 8 and 9 show plots of effective coupling versus energy in the minimal nonsupersymmetric and supersymmetric SU(5) models, extrapolated from the latest data.

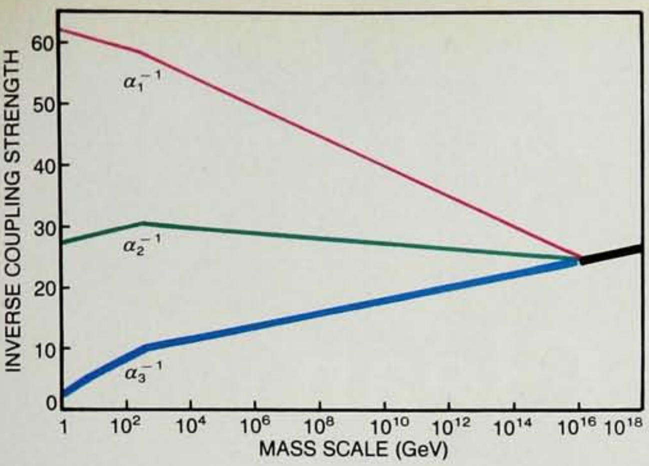
Together with the previous indications from proton decay, these new results provide highly suggestive, if circumstantial, evidence for virtual supersymmetry. They also greatly reinforce the case for color unification. A minimal supersymmetric model is certainly not the only way to reconcile the existing data with color unification. More complex unification schemes, typically involving new particles with exotic quantum numbers and more complicated symmetry-breaking patterns, are also contenders.[22] At the moment these other contenders seem less compelling than the minimal supersymmetric model.

Prospects

If we take these indications of unification of couplings and virtual supersymmetry at face value, they both brilliantly confirm old ideas in particle theory and augur a bright future for particle experimentation. Within the next year or so the electron–proton collider HERA should be gathering data that will both test QCD and refine the determination of the strong coupling constant, whose uncertainties are currently the most important factor limiting comparison of theory and experiment. If virtual supersymmetry is operative well below the unification scale, Nature would be perverse not to use it in addressing the gauge hierarchy problem. If Nature is not perverse in this way, the masses of the superpartners cannot be too large, and real supersymmetry should not elude the next generation of acclerators (CERN's Large Hadron Collider or the SSC).

There is also good news regarding the search for proton decay. Taken at face value, the best fits of minimal superunified models to the couplings predict a unification scale of about 10^{16} GeV and a proton lifetime of about 10^{33} years through gauge boson exchange.[21] This lifetime is slightly outside the reach of existing experiments, but not hopelessly so. In supersymmetric models there are additional mechanisms for proton decay—involving decay into virtual scalar quarks—that do not occur in nonsupersymmetric models.[23] The rate of decay through these modes depends on details of aspects of the models that are poorly understood, and so it cannot be predicted with precision. However, in a wide class of models these modes dominate the decay and lead to extremely unusual final states.[24] Thus if the proton is ever observed to decay, the nature of its decay modes may give strong clues as to the nature of the unified theory underlying its demise.

Finally, we would like to make some simple observations about how gravity might fit into this picture. The coupling of the graviton may also be considered to run, and much faster than the other couplings (see figure 6). Because it is characterized by a dimensional coupling—Newton's constant—rather than the dimensionless couplings that characterize the other interactions, it increases (to a first approximation) linearly with energy, rather than logarithmically. However, it starts out so small that its extrapolation only meets the other couplings at approximately 10^{19} GeV, the Planck mass. This is comparable to, but definitely greater than, the unification scale. An important implication of this is that gravitational corrections do not drastically affect the running of couplings at or below the unification scale, on which the preceding discussion was based. The ratio of these scales—the Planck mass and the grand unification scale—is another fundamental dimensionless number, whose calculation presents an inspiring challenge to theoretical physics.



Minimal supersymmetric SU(5) model does cause the couplings to meet in a point. While there are other ways to accommodate the data, this straightforward, unforced fit is encouraging for the idea of supersymmetric grand unification. (Adapted from a figure provided by Amaldi.) **Figure 9**

We wish to thank Ugo Amaldi, Wim de Boer, John Ellis, Hermann Fürstenau and Luis Ibanez for valuable discussions.

References

1. Other good semipopular introductions to many of the main ideas of gauge theories include S. Weinberg, Sci. Am., July 1974, p. 50; and G. 't Hooft, Sci. Am., June 1980, p. 104. Standard textbooks on gauge theories include I. Aitchison, A. Hey, *Gauge Theories in Particle Physics*, Adam Hilger, Bristol, UK (1982); K. Gottfried, V. Weisskopf, *Concepts for Particle Physics I* and *II*, Clarendon, Oxford (1984); C. Quigg, *Gauge Theories for the Strong, Weak, and Electromagnetic Interactions*, Benjamin/Cummings, Reading, Mass. (1987); and T. P. Cheng, L.-F. Li, *Gauge Theories of Elementary Particle Physics*, Clarendon, Oxford (1984). See also the excellent annotated reprint collection *Gauge Invariance*, T. P. Cheng, L.-F. Li, eds., Am. Assoc. Phys. Teachers, College Park, Md. (1990).
2. An early attempt at unification of couplings is J. Pati, A. Salam, Phys. Rev. D **8**, 1240 (1973). Color unification along the lines discussed here was introduced in H. Georgi, S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).
3. R. Becker-Szendy *et al.*, Phys. Rev. D **42**, 2974 (1990). Particle Data Group, Phys. Lett. B **239**, 1 (1990).
4. For discussion of the Higgs phenomenon in quantum field theory, see Y. Nambu, Phys. Rev. **117**, 648 (1960); P. W. Higgs, Phys. Rev. Lett. **12**, 132 (1964); *ibid.*, **13**, 508 (1964); F. Englert, R. Brout, Phys. Rev. Lett. **13**, 321 (1964); G. S. Guralnik, C. R. Hagen, T. W. B. Kibble, Phys. Rev. Lett. **13**, 585 (1964). For the Higgs phenomenon in superconductivity, see F. London, H. London, Proc. R. Soc. London, Ser. A **149**, 71 (1935); V. L. Ginzburg, L. D. Landau, Zh. Eksp. Teor. Fiz. **20**, 1064 (1950); P. W. Anderson, Phys. Rev. **110**, 827 (1958); *ibid.*, **112**, 1900 (1958).
5. D. J. Gross, F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973). H. D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973).
6. D. J. Gross, F. Wilczek, Phys. Rev. D **8**, 3633 (1973). S. Weinberg, Phys. Rev. Lett. **31**, 494 (1973). H. Fritzsch, M. Gell-Mann, H. Leutweyler, Phys. Lett. B **47**, 365 (1973).
7. G. Miller *et al.*, Phys. Rev. D **5**, 528 (1972). A. Bodek *et al.*, Phys. Rev. D **20**, 1471 (1979).
8. R. P. Feynman, Phys. Rev. Lett. **23**, 1415 (1969). J. D. Bjorken, Phys. Rev. **178**, 1547 (1969).
9. For a recent review, see G. Altarelli, Ann. Rev. Nucl. Part. Phys. **39**, 357 (1984).
10. R. J. Hughes, Phys. Lett. B **97**, 246 (1980); Nucl. Phys. B **186**, 376 (1981). N. K. Nielsen, Am. J. Phys. **49**, 1171 (1981).
11. H. Georgi, H. Quinn, S. Weinberg, Phys. Rev. Lett. **33**, 451 (1974).
12. Yu. Gol'fond, E. Likhtman, JETP Lett. **13**, 323 (1971). D. Volkov, V. Akulov, Phys. Lett. B **46**, 109 (1973). J. Wess, B. Zumino, Phys. Lett. B **49**, 52 (1974). See also the very useful reprint collection *Supersymmetry* (2 vols.), S. Ferrara, ed., North-Holland/World Scientific, Singapore (1987).
13. Y. Nambu, Phys. Rev. Lett. **4**, 380 (1960). Y. Nambu, G. Jona-Lasinio, Phys. Rev. **122**, 345 (1961); **124**, 264 (1961). J. Goldstone, Nuovo Cimento **18**, 154 (1961). P. Fayet, J. Iliopoulos, Phys. Lett. B **51**, 461 (1974). L. O'Raifeartaigh, Nucl. Phys. B **96**, 331 (1975).
14. Early papers mentioning the hierarchy problem include E. Gildener, Phys. Rev. D **14**, 1667 (1976); E. Gildener, S. Weinberg, Phys. Rev. D **15**, 3333 (1976); L. Susskind, Phys. Rev. D**20**, 2619 (1979). Early papers applying supersymmetry in atempts to ameliorate the hierarchy problem include M. Veltman, Acta Phys. Pol. B **12**, 437 (1981); S. Dimopoulos, S. Raby, Nucl. Phys. B **192**, 353 (1981); M. Dine, W. Fischler, M. Srednicki, Nucl. Phys. B **189**, 575 (1981); E. Witten, Nucl. Phys. B **188**, 513 (1981); S. Dimopoulos, F. Wilczek, in *Unity of the Fundamental Interactions*, A. Zichichi, ed., Plenum, New York (1983), p. 237.
15. S. Dimopoulos, S. Raby, F. Wilczek, Phys. Rev. D **24**, 1681 (1981).
16. D. Freedman, P. van Nieuwenhuizen, S. Ferrara, Phys. Rev. D **13**, 3214 (1976). S. Deser, B. Zumino, Phys. Lett. B **62**, 335 (1976). See also Ferrara, ref. 12.
17. P. Fayet, Phys. Lett. B **69**, 489 (1977); **84**, 416 (1979). S. Ferrara, L. Ghirardello, F. Palumbo, Phys. Rev. D **20**, 403 (1979).
18. S. Dimopoulos, H. Georgi, Nucl. Phys. B **193**, 150 (1981).
19. N. Sakai, Z. Phys. C **11**, 153 (1981). L. E. Ibanez, G. G. Ross, Phys. Lett. B **105**, 439 (1981). M. B. Einhorn, D. R. T. Jones, Nucl. Phys. B **196**, 475 (1982). W. J. Marciano, G. Senjanovic, Phys. Rev. D **25**, 3092 (1982).
20. These results from the DELPHI, ALEPH, L3 and OPAL collaborations are in a series of papers too numerous to list here; they mostly have appeared (and continue to appear) in Phy. Lett. B.
21. U. Amaldi, W. de Boer, H. Fürstenau, Phys. Lett. B **260**, 447 (1991). The possibility of such analysis was demonstrated earlier by J. Ellis, S. Kelly, D. Nanopoulos, Phys. Lett. B **249**, 441 (1990). A similar analysis was carried out independently by P. Langacker, M. Luo, U. Pennsylvania preprint (1991).
22. See, for example, H. Georgi, D. Nanopoulos, Nucl. Phys. B **155**, 52 (1979); L. E. Ibanez, Nucl. Phys. B **181**, (1981); P. Frampton, S. L. Glashow, Phys. Lett. B **135**, 340 (1983); P. Frampton, B.-H. Lee, Phys. Rev. Lett. **64**, 619 (1990); A. Giveon, L. Hall, U. Sarid, U. Calif., Berkeley preprint (July 1991).
23. S. Weinberg, Phys. Rev. D 26, 257 (1982). N. Sakai, T. Yanagida, Nucl. Phys. B **197**, 533 (1982).
24. S. Dimopoulos, S. Raby, F. Wilczek, Phys. Lett. B **112**, 133 (1982). J. Ellis, D. V. Nanopoulos, S. Rudaz, Nucl. Phys. B **202**, 43 (1982). ■