ON THE UNIFICATION OF PHYSICS

In the past forty years, theoretical physics has undergone a transformation in its avowed objectives as radical as any that has occurred in the last three centuries. This redirection has been at once tremendously effective and mysteriously quiet, a sort of velvet revolution in the conception of the aim of physical theory. Unlike scientific revolutions that commonly preoccupy philosophers of science, this change has not come in accompanied by the battle cries of opposing camps, nor been punctuated by any dazzling explosion of empirical evidence. Indeed, this change has not been, in the common usage of the term, a scientific revolution at all, which perhaps explains why it has progressed so seemingly naturally and inexorably, like the advance of the seasons. The transformation centers around the remarkable idea that the aim of physical theory is to achieve unification.

Today, anyone inquiring, whether in popular or professional literature, into the current status of fundamental physical theory is virtually guaranteed to be told the following tale. In the first part of this century, physicists had verified the existence of four basic physical forces: electromagnetism, gravity, the strong nuclear force, and the weak nuclear force. Passably accurate theories of these forces individually have been developed, but those theories do not yet demonstrate any deep connection among all of the forces. The aim of physics is now to produce theories which unify these forces, which show, ultimately, that there is at base only one fundamental force in the universe, which has come to display itself as if it were many different forces.

The first step in this program has already been taken: electromagnetism has been unified with the weak nuclear force in the electroweak theory. The other steps, though still to be achieved, have already been named, and are to occur in a particular sequence. The electroweak force is to be unified with the strong nuclear force by a *grand unified theory* (GUT), and then, in the final step, the GUT will somehow be unified with gravity in a *theory of everything* (TOE).

This image of the future course of physical theory has become so pervasive as to rank almost as dogma. Still, as mentioned above, the process by which it has become so widely accepted would not generally be regarded as any sort of scientific revolution. For what has been accepted is not itself a theory, and could not be defended or criticized as an empirical theory would be. It is instead a commit-

0022-362X/96/9303/129–44

ment (that might be deep or shallow) to a general view about how theory is likely to develop and about the ultimate nature of as yet undiscovered laws. It is certainly a commitment that is not so strong as to override empirical support for theories that do not follow the projected path. But, nonetheless, it is a powerful image that helps shape the direction of research, and philosophers of science ought to be intrigued about how physicists have come to find themselves in the unification business.

The notion that all fundamental forces are somehow deeply unified goes back at least to Isaac Newton. Albert Einstein, of course, spent the latter years of his life searching for a unified field theory, but one along lines quite different from those envisaged today. So our object of inquiry is not the idea of unification in all of its generality. Rather, we want to explore the particular species of unification foreseen in the scenario sketched above. Of this contemporary brand of unification we want to ask two questions. First is the fundamental philosophical question: What is it? What exactly is intended by "unification" in these contexts? Second, and more contentiously, we must ask: Why think that these forces are unified in the manner envisaged? Unification puts rather strong constraints on the form of a physical theory, and it is surely appropriate to ask what grounds we have to believe that successful theories will respect those constraints. By what modern Parmenidean logic have so many contemporary physicists come to the conclusion that "All is One"? Or more directly, is there any *empirical* ground for faith in the project of unification, and if so, how strong is it?

But first things first: we shall begin by asking just what unification amounts to. As is common in rough initial surveys, we begin by staking out some extreme boundaries.

I. WHAT UNIFICATION IS NOT

Let me start with the obvious. It is a universally accepted desideratum that theories of the various forces be *consistent* with one another, but consistency is clearly not sufficient for unification. Indeed, we want all accepted scientific theories from all domains to be consistent with one another. But the fact that a theory of embryonic development does not contradict a theory of the formation of the rings of Saturn is surely insufficient to render the two unified.

Of course, acknowledging consistency as a goal does not render it trivial or easy to achieve. It is not entirely clear if quantum theory is consistent with general relativity, and it might turn out that demanding mere consistency of our physical theories severely limits the form they can take. Possibly, one might even suspect that the only way to render theories of the fundamental forces consistent is to unify them. But it is hard to imagine a convincing argument for such a position, and in any case, unification itself is a stronger condition than mere consistency. What more is being demanded?

A stronger condition than consistency is the employment of a single fundamental dynamics, but this too falls short of unification. Consider, for example, the theory one gets by conjoining Newtonian dynamics with the Law of Universal Gravitation and with Coulomb's Law. In this case, the theories of gravity and of the electrical force are not so disjoint as, say, child psychology and fluid dynamics. The accounts one would give of electrical and gravitational effects would share a common explanatory structure. But even so, no one would regard this as a case of unifying gravity and electricity. Although both theories employ a common dynamical theory, still the forces are not postulated to have anything in particular to do with one another. One could, for example, model a world with gravitational but no electrical forces. Furthermore, the presence or absence of one force would have no bearing on the presence or absence of the other. Unification, then, must be supposed to go beyond mere commonalty of dynamics.

The next step up is law-like connection or correlation among physical forces. The paradigm here would be Maxwell's theory of electromagnetism, in which variation in certain electrical quantities gives rise to magnetic phenomena, and vice versa. Unlike the case of gravity and the electric force, neither electricity nor magnetism can be understood without reference to the other. Indeed, in some sense, Maxwell's theory does unify electricity and magnetism. But in another, deeper sense, electric and magnetic fields retain a completely distinct ontological status in Maxwell's theory. They may be nomically correlated, they may give rise to one another, but at base they are still entirely different entities.

Indeed, the failure of classical electromagnetic theory to unify electric and magnetic phenomena was the leading complaint voiced in Einstein's[1] "special relativity" paper:

> It is known that Maxwell's electrodynamics—as usually understood at the present time—when applied to moving bodies, leads to asymmetries which do not appear to be inherent in the phenomena. Take, for example, the reciprocal electrodynamic action of a magnet and a conductor. The observable phenomenon here depends only on the rela-

[1] "On the Electrodynamics of Moving Bodies," reprinted in *The Principle of Relativity* (New York: Dover, 1952).

> tive motion of the conductor and the magnet, whereas the customary view draws a sharp distinction between the two cases in which either one or the other of these bodies is in motion. For if the magnet is in motion and the conductor at rest, there arises in the neighbourhood of the magnet an electric field with a certain definite energy, producing a current at the places where parts of the conductor are situated. But if the magnet is stationary and the conductor in motion, no electric field arises in the neighbourhood of the magnet. In the conductor, however, we find an electromotive force, to which in itself there is no corresponding energy, but which gives rise—assuming equality of relative motion in the two cases discussed—to electric currents of the same path and intensity as those produced by the electric forces in the former case (*ibid.*, p. 37).

The *failure* of Maxwell's theory to unify electricity and magnetism, that is, to show that in the two cases described one really has *identical* physical situations, led to the Special Theory of Relativity (STR). And it is this deeper sense of unification, the idea that all the physical forces are at base one and the same, which contemporary physicists invoke when they speculate on the theories to come.

We now have a lower limit in our search for the meaning of unification. Consistency, common dynamics, and nomic correlation are all features we might seek when constructing theories of forces, but they all fall short of unification. We shall now encircle our quarry by describing some cases of perfect unification, thus setting an upper limit that cannot be surpassed. As we shall see, between the upper and the lower bounds, several different levels of unification will become discernible.

II. PERFECT UNIFICATION: TWO EXAMPLES

Since Einstein complained that Maxwell's theory failed to unify electrical and magnetic phenomena, the first place to look for successful unification is STR. And in that theory, the formerly distinct electric and magnetic fields are so commingled that a more complete integration is impossible to imagine. Indeed, the very words 'commingled' and 'integration' are inappropriate here, implying, as they do, two things that are being somehow combined. But in STR there is truly but one thing: the electromagnetic field tensor. It is not that the electric field is reduced to the magnetic, but that both are shown to be merely frame-dependent artifacts, inessential and misleading ways of describing the single objective reality. So STR resolves the problem of inductive currents that Einstein describes in this way: when a current flows in the conductor, there is an electromagnetic field tensor in the vicinity of the conductor. There is simply no objective fact of the matter about whether or not there is an electric *field* near the conductor, or what the value of the magnetic field is. Electric and magnetic fields *are not objectively real*, they "arise" only when one chooses a certain reference frame relative to which the phenomena are to be described. Thus, the electric and magnetic fields are "unified" by being, in a way, eliminated entirely from the fundamental ontology, and by being replaced by a single, frame-independent entity. To paraphrase H. Minkowski's famous remark on space and time in special relativity, henceforth the electric field by itself and the magnetic field by itself are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.

A second example of fundamental unification is provided by the second Einsteinian revolution: the General Theory of Relativity (GTR). In the Newtonian milieu, gravitational and inertial structure are quite different. In anachronistic terminology, inertial structure is just the affine structure of Newtonian space-time. Newton's first law states that an object with no forces on it will travel along a straight trajectory through space-time. Gravity, on the other hand, is a force that deflects objects from their inertial paths. In Newtonian mechanics, gravity has no more intimate connection to inertia than does electricity, and the equality of inertial mass and gravitational mass would be no more expected a priori than the equality of electric charge and inertial mass.

In the general theory, gravity and inertia are reduced to a single structure: the metrical structure of space-time. One may retain Newton's first law, but only if one recognizes that *there is no force of gravity at all*. Phenomena formerly understood as effects of gravitational forces are now explained as effects of the influence of matter on the affine structure of space-time. The equality of inertial and gravitational mass, as evidenced by free fall in a gravitational field, is reinterpreted as the common response of all matter to inertial structure. Objects do not couple to the gravitational field, they merely exist in space-time.

Unlike the case of electromagnetic unification, one is perhaps inclined to regard the unification here as a *reduction* of gravity to inertia, especially since Newton's Law of Inertia still holds (in some sense) while the Law of Universal Gravitation does not. But the dispute is a minor one. Newtonian theory has two distinct entities: Newtonian (or Neo-Newtonian) space-time and the gravitational field. GTR has but one: curved space-time. Neither of the original two entities survives unscathed in the later theory (as happens in a true reduction), but rather both are replaced by a single new object.

Were the projected unification of electromagnetism, gravity, and the strong and the weak nuclear forces to run as deep as the examples we have just examined, then the very distinction among the forces would disappear. If we take the special relativistic account of electromagnetism as a model, a unified TOE should have the consequence that no effect could be objectively ascribed to the action of the weak nuclear force rather than gravity. It is difficult to envisage exactly how this could be so, but it seems fair to say that a deeper or more complete unification cannot even be described. We can safely allow these examples to serve as our paragons.

III. EVIDENCE FOR UNIFICATION

The two examples of perfect unification, relativistic electrodynamics and general relativity, also provide particularly clear examples of one sort of answer to our second question: Why think that the laws of nature ought to be unified at all? In each of these cases, the unification serves to explain a manifest symmetry in phenomena that was known and remarked before the theories were developed. We have already mentioned both of these symmetries.

There is the invariance of the predictions of Maxwell's theory when one changes inertial frames, as exemplified in the fact that inductive effects depend only on the relative velocities of the magnet and the conductor. As Einstein noted, the fact that the magnitude of the effect depends only on the relative velocities stands in sharp contrast to the theoretical explanations given of the phenomenon according to Maxwell's equations. If the electric and magnetic fields are objective, frame-independent entities, then superficially identical experiments will receive deeply divergent explanations. Einstein touts the ability of his theory to eliminate this asymmetry of explanation as one of its main theoretical virtues.

For general relativity, the manifest symmetry was obviously the equality of inertial and gravitational mass. Although this was not so much a major motivation for developing the theory, still in retrospect we can now recognize it as a clue to the correct form of the theory of gravitation. It was certainly a fact that recommended itself to the attention of physicists, enough to inspire R. von Eötvös to perform his experiments long before the advent even of special relativity.

Note that although the equality of inertial and gravitational mass is afforded a satisfying explanation by GTR, it would be difficult to argue that the equality demanded the sort of unification that GTR provides. Newtonian theory can easily enough provide an explanation: it is not that there are really two kinds of mass that happen to turn out equal, but just that the single quantity, mass, happens to figure in two different laws. One and the same mass that endows an object with its resistance to acceleration also causes (and responds to) gravitational forces. Such an explanation would not, of course, fundamentally unify inertia and gravitational forces in anything like the thoroughgoing way that is achieved in GTR.

It should also perhaps be noted that GTR itself also supposes, at least in theory, a dual role for mass. One still calculates the mass of an electron or proton by its inertial effects: how much it accelerates in a given field. But, in principle, this same inertial mass contributes to the calculation of the stress-energy tensor, and so to the gravitational effect of the particle. I say 'in principle', since in practice the masses of astronomical bodies are determined from their gravitational effects. That is, one calculates the mass of the sun, for example, by determining the orbits of the planets and setting the mass to be the value demanded by the gravitational field equations to account for those orbits. Still, one would hope eventually to derive the stress-energy tensor from a theory of the composition of the star, and the masses derived from inertial effects would play a role, in such a determination. Here, presumably, one would again have one and the same quantity entering into very different fundamental laws.

In any case, GTR does explain the fact that small test objects under the influence only of the gravitational field follow parallel trajectories, and does this without postulating the equality of inertial and gravitational mass. To this extent it constitutes an advance over Newtonian theory, which can only get the result if the two "sorts" of mass are equal.

Our second question, then, can be stated best by use of an analogy. As the symmetry of induction effects suggested a deep relationship between the electric and magnetic fields, and as the laws of free fall suggested a deep connection between inertial and gravitational structure, do any manifest phenomena suggest that the electroweak force ought to be unified with the strong force, or either of these with gravitation? Of course, to answer this question properly we must get clear on exactly the sort of unification envisaged. Our clearest guide here is to be found in the example of electroweak unification, so I turn next to this case.

IV. UNIFICATION IN GAUGE THEORIES

With our map of varieties of unification in hand, we can now turn to the fundamental theoretical structure employed in current attempts at unification: gauge theories. A short reminder of the structure of gauge theories is in order.

Any gauge theory is based on a gauge group. The theory is constructed by first choosing a gauge group, such as U(1) for electromagnetism (the group of phase shifts $e^{i\theta}$) or SU(3) for chromodynamics. A Lagrangian is then constructed which is invariant under the group operations. The gauge particles, or carriers of the force, are associated with the generators of the group. Thus, U(1) has a single generator and a single gauge particle, the photon. SU(3) has eight generators, yielding the eight sorts of gluons. Particles are then assigned to multiplets that form representations of the group. That is, the members of the multiplet are transformed into one another by the action of the group operators. Part of the game is to find multiplets of the right size and physical properties to contain all of the known particles that feel the force.

There is an interesting complication to the structure sketched above. In order for the Lagrangian to be invariant under the group operations, the gauge particles must be massless. It was thought for some time that this necessarily implied that the forces modeled must be long distance, like electromagnetism or gravity. If so, then the whole machinery would be inappropriate for the representation of short-range forces, such as the weak and strong nuclear forces. One could insert masses for the gauge particles into the equations by hand, but doing so would break the gauge invariance (and also seemed to render the equations nonrenormalizable). These problems were finally resolved by P. W. Higgs, who demonstrated a mechanism by which a Lagrangian that contains only massless gauge particles can give rise to a phenomenology that contains only massive particles. This is accomplished through spontaneous symmetry breaking.

Where, in all of this, are we to seek the unification of forces? Since the forces are generated out of the gauge group, we should begin there. Various sorts of groups have been proposed, but it is best to start with the standard model.

The standard model employs the gauge group SU(3) x SU(2) x U(1), that is, it employs a product group rather than a simple group. Product groups are constructed from the simple groups by a simple combinatorial method. Thus, suppose I have a disk that can be rotated about its center. The symmetry group is U(1) (like the phase of the wave function). A sphere can be rotated about any axis in three-dimensional space, so the symmetry group is SO(3). Now, if I think of the compound object consisting of the disk and the sphere as a single entity whose states are specified by giving the orientation of the disk and the orientation of the sphere, then the set of operations that I can perform on the compound object constitutes the group U(1) x SO(3). That is, the set of ordered pairs that consist of one operation from U(1) and another from SO(3) itself obviously forms a group, whose group structure is U(1) x SO(3). Any two groups can be combined in this way, even if they have no interesting structure in common.

In part, this is how the standard model was constructed. Chromodynamics was constructed using SU(3) while the electroweak theory used SU(2) x U(1). These theories have no intrinsic relation to one another, besides both being gauge theories. Hence the standard model was constructed by simply pasting the two groups together.

The strong and electroweak theories are no more intrinsically unified in the standard model than Newtonian gravity and the Coulomb force are in Newtonian mechanics. In the taxonomy developed above, this is a case of common dynamics and nothing more. This is, of course, as it should be, for if the standard model were deeply unified, there would be no call for GUTs.

What of the electroweak theory? The gauge group here is SU(2) x U(1). Since the group of quantum electrodynamics (QED) was just U(1), this looks on the surface no better than the combination of the strong and electroweak forces, with the U(1) group accounting for electromagnetism and the SU(2) for the weak force. But the situation is a bit more complicated.

There are three generators of the SU(2) symmetry. Two of these are associated with gauge particles in the usual way: the W^+ and W^-. The third generator ought to be associated with a neutral particle, which we shall call the X^0 for the moment. The U(1) symmetry is also associated with a neutral gauge particle, which we may call Y^0. For purely empirical reasons, one cannot identify the X^0 with the third physical particle involved in weak interactions (namely, the Z^0), nor the Y^0 with the photon. Instead, in order to get the multiplets to come out right, one must identify both the photon and the Z^0 with *mixtures* of the X^0 and Y^0. The precise proportion of X^0 and Y^0 that go to make up the photon and the Z^0 is given by the so-called *mixing angle*, which is a free parameter in the standard electroweak theory. It is in this mixing, and in it alone, that electromagnetism and the weak force become unified in the electroweak theory. One simply cannot write down an adequate theory of the weak interaction without also including the materials for the electromagnetic one. (The converse, interestingly, is not true, as is demonstrated by the existence of QED. The reason for this is that a combination of a particular gauge transformation from the SU(2) group with one from the

U(1) group has the effect of only changing the phases of charged particles.)

To what extent is the mixing involved in the electroweak theory really a case of unification? It is at least as strong as the unification of electricity and magnetism in Maxwell's theory, in that the equations describing the weak force must also describe electromagnetism. And there is something a bit deeper, since the photon and the Z^0, the observed neutral particles associated with the weak and electromagnetic forces, are both built from the X^0 and Y^0. Still, the unification fails to reach the level of perfection found in GTR, or even the level anticipated in the GUTs. Thus, we find some ambivalence among physicists about how exactly to describe the unification involved. K. Moriyasu[2] states:

> The weak and electromagnetic gauge fields are now completely unified. What is most interesting about the unification is the mixing of the U(1) and SU(2) gauge fields that was necessary to construct the physical electromagnetic potential. We began with a product of disconnected symmetry groups and ended up by unifying them through a mixing of gauge fields. The reason for the mixing, of course, has nothing to do with gauge theory *per se.* It was built in "by hand" through the identification of the leptons as the appropriate doublets and singlets of weak-isotopic spin (*ibid.*, pp. 109–10).

H. M. Georgi[3] puts it this way:

> The SU(2) x U(1) theory is not particularly beautiful. It is often called a unification of the weak and electromagnetic interactions, but, in fact, the unification is partial at best. The problem is the U(1) charge.... [T]his is a charge that commutes with all the other weak and colour charges, so group theory tells us nothing about it. In particular, because of the U(1), the theory gives us no explanation of the striking fact of electric charge quantization. Further, because of the U(1), there are two separate dimensional charges or coupling constants required to specify the theory, one for the SU(2) charge e_2 and coupling constant α_2, and another for the U(1) charge e_1 and coupling constant α_1. This introduces another unknown parameter into the theory and again reduces its explanatory power (*ibid.*, p. 437).

Thus, for Georgi in particular, the fact that the group structure of the electroweak theory is a product group still renders the theory unsatisfactorily disunified, despite the mixing.

[2] *An Elementary Primer for Gauge Theory* (Singapore: World Scientific, 1983).

[3] "Grand Unified Theories," in P. Davies, ed., *The New Physics* (New York: Cambridge, 1989).

There is, then, an even higher level of unification to be sought in gauge theory, namely, when the gauge group is a simple group. This is the aim of the GUTs: to produce a gauge theory with a single, uncompounded gauge group from which the strong, weak, and electromagnetic forces may all be derived. The simplest such theory uses the group SU(5), but other options are available.

So in moving to gauge theories, we have found three levels of structure. Any two independently accepted gauge theories can be pasted together with a product group to get a theory that nominally has one gauge group. This does not constitute any sort of unification of the theories at all, and corresponds to having two distinct forces at play in theories that share the same basic dynamics. At the second level, such a product group, such as SU(2) x U(1), can give rise to physically observable forces whose gauge particles derive from mixing of the groups. It is questionable whether one should regard this as any sort of deep unification. It seems closest to cases, such as Maxwell's theory, in which different fields become coupled by the equations, except that the *underlying* gauge groups only couple if one demands that the observed particles be generated. (Question: at high energies, when the theory becomes completely symmetrical again, would the U(1) and the SU(2) simply decouple?) At the third level are gauge field theories premised on a simple gauge group. This is the sort of unification sought by the GUTs.

In what sense would a theory of the this sort achieve a unification of the forces? Certainly in the sense that the various forces would all ultimately derive from a single underlying structure. But there is a deeper unification than just this. The Lagrangian of that underlying theory is invariant under the gauge transformations. In that sense, the basic physics does not recognize any *generic* difference among the forces, just as the rotation invariance of a Lagrangian would demonstrate that the physics did not postulate any generic difference among the three dimensions of space. This is, of course, not unification in the sense of reduction to unity: space is still three dimensional. The three dimensions, though, are not intrinsically different, and, indeed, it is physically arbitrary how one divides the three-dimensional object into three directions.

Such thorough unification had best not show up in all contexts, since the world we deal with is manifestly not completely invariant in this way. That is, electromagnetic, weak, and strong interactions are clearly distinguishable, and cannot be transformed into one another by any simple change of reference frame in the way electric and magnetic fields can in Maxwell's theory. The symmetries in the

Lagrangian are to be spontaneously broken by the Higgs mechanism, and would be manifest only at very high temperatures. Spontaneous symmetry breaking is what hides the deep unification of the forces from our eyes. (Spontaneous symmetry breaking is also needed to render the theory renormalizable.)

The typical analogy used for spontaneous symmetry breaking is ferromagnetism: even though the fundamental laws are invariant under spatial rotations, the physical object governed by those laws may not be, and in a particularly salient way. Furthermore, if the laws include some stochastic component, one can even have a situation in which an invariant initial state evolves by means of invariant laws to a state that breaks the symmetry. (The usual examples of crystal formation in freezing are a bit misleading in this way, since in those cases one could imagine that an initial state is not really isotropic, though it may be in its macroscopic variables.) And for a certain use, this analogy is perfectly adequate.

But in one sense, the analogy badly fails. In the case of the ferromagnet, one does not add any new physics in the process of symmetry breaking: the interactions among the particles of the magnet do the job. But the symmetry of the gauge theories is different. In the electroweak theory, the SU(2) x U(1) symmetry does not just evaporate of its own accord. To get the symmetry to break (and to give one of the massless gauge particles a mass) one needs to add a new bit of physics: the Higgs field. The field does not per se break the symmetry (so once one has the gauge field plus the Higgs field, the analogy with ferromagnetism can proceed) but still from the point of view of unification, the addition of the new field must come as a disappointment. We manage to unify electromagnetism and the weak force, say, in a single gauge field, but only by postulating yet another (heretofore unsuspected) scalar field.

Is the scalar field itself unified with the gauge field in any interesting way? As in the case of the unification of SU(2) x U(1) in electroweak theory, the answer seems to be that the fields are only unified in that they all cooperate to produce the particular (apparently diverse) forces and particles we see. At the very high temperatures at which those forces would unify, the gauge field and the Higgs field would decouple. Thus, we are faced with a new question pertaining to unification: Should we expect or demand any GUT to unify the gauge and scalar fields in any way? This completes our survey of the meaning of "unification" in unified theories of various forces.

V. EMPIRICAL CONSIDERATIONS

If the preceding account is correct, we now have some idea of the sort of unification of forces sought in the GUTs, and presumably ultimately in a TOE. This leaves us with our second question: What reason, if any, is there to believe that the world really is unified in this way? That is, before a unified theory has actually been developed, are there empirical clues or indications akin to the symmetry of electromagnetic induction effects or the equality of inertial and gravitational mass which seem to point to a deep unification of the forces?

Obviously, manifest symmetry of the forces is precluded by the spontaneous symmetry breaking. Unification is to be sought in spite, rather than because, of the immediately observable properties of the forces. The mechanism of symmetry breaking allows the research program to continue in the face of the apparent dissimilarity of the forces, but it also denies us direct empirical grounds for believing that there is any hidden symmetry at all.

Unification might be sought for purely aesthetic reasons, and one occasionally finds sentiments of the "all is one" variety expressed in the literature. But this is surely the thinnest of all possible reeds on which to found a research program.

Unification might be sought on the general methodological grounds of repeating strategies that succeeded in the past. The line would be that the successful elucidation of the weak force ultimately demanded incorporation of electromagnetism, so perhaps the successful elucidation of the strong force should also require unification with the electroweak. But this argument fails on several grounds. First, unification was not a strategy that was followed in arriving at the electroweak theory; rather, the unification was forced on those who were primarily engaged in seeking an adequate theory of the weak force. Second, the sort of unification that led to success in the electroweak theory is not the sort of unification sought in the GUTs. As we have seen, some theorists deny that the electroweak theory displays any real unification of electromagnetism and the weak force at all. Third, the situation vis-à-vis the weak and strong forces is quite different. No workable theory of the weak force existed before the unified theory, but quantum chromodynamics (QCD) does exist, and seems to work. (Incidentally, this is one way in which the usual story about unifying forces is wrong. It is not that at some point we had *theories* of the electromagnetic, weak, strong, and gravitational forces separately, and now we have managed to unify the first two. Rather, at some point we *recognized the existence* of all four forces, and found that unification was needed to account for the weak force.) So why not think that SU(3) x SU(2) x U(1) is all there is?

Here is what Georgi has to say about the motivation for SU(5):

> Once we understood SU(2) x U(1) and quantum chromodynamics, Grand Unified Theories were a simple step. The motivation for the simplest GUT, SU(5), was not any mystical desire to follow in Einstein's footsteps and unify everything. Shelley Glashow and I were just trying to understand SU(2) x U(1) better. For several years, we had realized that if we could incorporate the SU(2) x U(1) gauge symmetry into a single simple group it would give us some extra information. It would fix the value of the weak mixing angle, a free parameter in the SU(2) x U(1) theory and it would explain why all the electric charges we see in the world are multiples of the charge of the electron (*op. cit.*, p. 454).

As Georgi notes, these considerations all stem from the desire to complete the only partially unified SU(2) x U(1) theory in a more satisfactory way, and demand only that that theory be derived from some simple group, not that the simple group also account for any other forces. Of course, the SU(5) theory did attempt to incorporate the strong force, but that was mostly due to the observation that quarks together with the electron and neutrinos could be fit into the multiplets of SU(5). In any case, the simplest SU(5) theory seems not to work.

There is another class of considerations that have been taken to point to the GUTs. If the SU(2) x U(1) structure can be derived by symmetry breaking from a simple group, then, as Georgi notes, the mixing angle will be calculable from first principles. Using SU(5) as the simple group, one estimates the mixing angle to be 0.20 ± 0.01, while the observed value is 0.230 ± 0.015. Steven Weinberg[4] remarks that the theoretical calculation "...is in reasonable agreement with experiment. This is just a single quantitative success, but it is enough to encourage us that there is something in these ideas" (*ibid.*, p. 201).

All of this leads us at best to a GUT. What about gravity? Do we have any empirical grounds to expect gravitation to be unified with the other forces in any significant way?

Gravity presents a problem very different from the unification of the strong and electroweak forces. While it is clear that SU(2) x U(1) is at least consistent with SU(3) QCD, it is not yet clear how to wed GTR with quantum theory in a single theoretical framework.

[4] *Dreams of a Final Theory* (New York: Pantheon, 1992).

Indeed, there seems to be a fundamental incompatibility (beyond problems of renormalization) between the basic approaches of gauge field theories and GTR. The field theories explain forces as due to the action of fields, ultimately via the exchange of virtual gauge particles. Objects couple to the field only if they interact through some charge that serves as a coupling constant. Uncharged particles would be unaffected by the field. But according to GTR, gravity simply is not a force. There is no field whose effect is to deflect appropriately charged particles from their inertial trajectories. Particles do not couple to the gravitational field, they simply exist in space-time. Gravity does not deflect particles from their inertial paths, as Newton thought, it determines their inertial paths.

The incompatibility of the gauge field theory approach with the heart of GTR is most clearly demonstrated by the fact that in some gauge theories of gravity, the equivalence principle fails: antimatter is subject to different gravitational effects than matter. So we are left with a very perplexing situation. Some of those pursuing TOEs are willing to assert that what appears to be a perfect symmetry (the equality of inertial and gravitational mass) really is not so, and they reject the satisfying explanation of that symmetry by the unification of inertial and gravitational structure, doing so in the name of a supposed underlying symmetry among other forces and gravity which is completely shattered in the phenomena we observe.

What empirical grounds are there for believing gravity to be unified with any other forces? Weinberg cites the fact that the unification energy predicted by some GUTs is only a few orders of magnitude below the Planck energy (*ibid.*, p. 203), hardly very compelling.

Perhaps I have read too much into the rhetoric of some presentations. Physicists want a theory of gravity that is compatible with the theories of other forces, and perhaps this is all that is intended by "unification" in this case. If more is demanded, we are within our rights to ask what form this deeper unification is supposed to take and what reasons we have to suspect that the world is unified in the envisaged way. At this point, there is little hard evidence for the kind of structure postulated by the GUTs and even less for the TOEs. But the last word should rest with the physicists, and I shall give it to Richard Feynman. When Robert Crease went to interview Feynman[5] on the history of the standard model, the following exchange took place:

[5] Recounted in James Gleick, *Genius: The Life and Science of Richard Feynman* (New York: Pantheon, 1992).

When a historian of particle physics pressed him on the question of unification in his Caltech office, he resisted. "Your career spans the period of the construction of the standard model," the interviewer said.

" 'The standard model,' " Feynman repeated dubiously.

"SU(3) X SU(2) X U(1). From renormalization to quantum electrodynamics to now?"

"The standard model, the standard model," Feynman said. "The standard model—is that the one that says that we have electrodynamics, we have weak interaction, and we have strong interaction? Okay. Yes."

The interviewer said, "That was quite an achievement, putting them together."

"They're not put together."

"Linked together in a single theoretical package?"

"No."

The interviewer was having trouble getting his question on the table. "What do you call SU(3) X SU(2) X U(1)?"

"Three theories," Feynman said. "Strong interactions, weak interactions, and electromagnetic.... The theories are linked because they seem to have similar characteristics.... Where does it go together? Only if you add some stuff that we don't know. There isn't any theory today that has SU(3) X SU(2) X U(1)—whatever the hell it is—that we know is right, that has any experimental check.... Now, these guys are trying to put all this together. They're *trying* to. But they haven't. Okay?"...

"So we aren't any closer to unification than we were in Einstein's time?" the historian asked.

Feynman grew angry. "It's a crazy question!...We're certainly closer. We know more. And if there's a finite amount to be known, we obviously must be closer to having the knowledge, okay? I don't know how to make this into a sensible question.... It's all so stupid. All these interviews are always so damned useless."

He rose from his desk and walked out the door and down the corridor, drumming his knuckles along the wall. The writer heard him shout, just before he disappeared: "It's goddamned useless to talk about these things! It's a complete waste of time! The history of these things is nonsense! You're trying to make something difficult and complicated out of something that's simple and beautiful."

Across the hall Murray Gell-Mann looked out of his office. "I see you've met Dick," he said (*ibid.*, pp. 433–34).

TIM MAUDLIN

Rutgers University