The 1979 Nobel Prize in Physics

This year the Nobel Prize in Physics was awarded to Sheldon Glashow, Abdus Salam, and Steven Weinberg. Glashow and Weinberg are professors of physics at Harvard University; Salam is professor of physics at Imperial College, London, and director of the International Centre for Theoretical Physics, Trieste. The award occasioned little surprise and much approval within the community of high-energy theoretical physicists, where the work of Glashow, Salam, and Weinberg has become widely recognized as major achievements of the field.

There is a popular model of a breakthrough in theoretical physics: A field of physics is afflicted with a serious contradiction. Many attempts are made to resolve the contradiction; finally, one succeeds. The solution involves deep insights and concepts previously thought to have little or nothing to do with the problem. It unifies old phenomena and predicts unexpected (but eventually observed) new ones. Finally, it generates new physics; the methods used are successfully extended beyond their original domain.

As I shall attempt to explain here, the work of Glashow, Salam, and Weinberg fits this model almost perfectly. I will begin with the contradiction, as it troubled us in the late 1950's: The weak interactions were not renormalizable. This sentence requires six paragraphs of explication.

On the scale of the distances and energies characteristic of nuclear physics, the fundamental forces of nature fall into four distinct classes. The weakest force is gravitation; indeed, it is so weak as to be of no relevance to nuclear phenomena. Next in strength is the (unimaginatively named) weak interaction, the force responsible for many radioactive nuclear decays. Somewhat stronger is the familiar electromagnetic force. Finally, strongest of all is the (again unimaginatively named) strong interaction, the force that holds the neutrons and protons together in the nucleus.

These forces differ in ways other than strength. Two of these are all that need concern us here. First, electromagnetism and gravity fall off slowly with separation, while the strong and weak interactions fall off rapidly, becoming essentially negligible for particles separated by more than a few nuclear diameters. Second, not all forces act on all particles; for example, electromagnetism does not directly act on electrically neutral particles, and the strong interaction does not directly act on what are called leptons (electrons, muons, neutrinos, and so on).

No one is able to compute exactly the effects of any of these forces, even in the simplest cases. However, there are approximation methods that can sometimes be used to give predictions of great accuracy. The only one of these methods we need consider here is perturbation theory, an approximation method based on expansion in successive powers of some small quantity characteristic of the process under consideration. For example, for most electromagnetic processes, the first approximation predicts effects proportional to 10^{-3}, the second approximation predicts corrections to the earlier predictions proportional to 10^{-6}, the third approximation predicts further corrections proportional to 10^{-9}, and so on, until whoever is doing the computation gets tired.

Or until disaster occurs. For it is possible for the computation to explode. To be more precise, if one writes down an arbitrary force law consistent with relativity and quantum mechanics, no problem ever occurs in the computation of the first approximation. However, it may well be that the second approximation turns out to be infinite. (To use the example of the preceding paragraph, infinity may emerge as an infinite multiple of 10^{-6}, but this is still infinity.) Even if this disaster does not occur on the second approximation, it may still occur in one further on in the series. Only a small class of theories give finite predictions for all observable phenomena to all orders in the perturbation theory expansion. These theories are called renormalizable. (The reason for this curious name is too long and irrelevant a story to be told here.)

Renormalizability is tricky to spot; potentially infinite quantities may cancel each other at the last moment, and unless you set up the computation in just the right way you are liable to miss the cancellation. The quantum theory of electromagnetism was around for nearly 20 years before it was found to be renormalizable; before then it had been thought to make evidently infinite, evidently meaningless predictions.

I can now return to the situation in the late 1950's. The electromagnetic interaction was renormalizable, and the predictions of perturbation theory for electromagnetic processes were in splendid agreement with experiment. The strong interaction did not involve any small parameter that could be used for perturbation theory. The simplest quantum generalization of the classical theory of gravitation was apparently nonrenormalizable, but the theory was so complicated and the effects in high-energy physics of such small magnitude that most researchers were happy to postpone consideration of this problem.

The weak interaction was a different story. The small quantity of perturbation theory was roughly 10^{-5}. The first approximation was big enough to be observable and the second approximation, proportional to 10^{-10}, should have been negligible. By the late 1950's, a theory of the weak interaction had been found such that the first approximation was in agreement with experiment wherever the theory could be tested. The second approximation was $10^{-10} \times$ infinity; the theory was not renormalizable.

There were two possible ways out. One was that perturbation theory was deceptive, that the fault lay in the approximation rather than in the theory. Much effort was devoted to this possibility, but nothing substantial resulted. The alternative was that the true theory of weak interactions was some undiscovered renormalizable theory that mimicked the predictions of the current theory in all experiments yet done.

In his 1958 Ph.D. thesis, Glashow suggested that such a theory would involve the unification of weak interactions and electromagnetism (*1*). There were tantalizing similarities between these two interactions. The electromagnetic interaction between two charged particles could be thought of as being mediated by the exchange of a quantum of light, a photon. A photon is one of a class of particles called vector mesons. The weak interactions could also be thought of as mediated by the exchange of a vector meson, called the W. The striking differences in range and strength between the two interactions could be explained by assuming the W was massive (on the order of 100 times more massive than the proton), in contrast to the massless photon. A unification along these lines had been advanced shortly before by Glashow's thesis adviser, Julian Schwinger (*2*). Glashow, however, was the first to connect the idea with renormalization.

By 1960, Glashow thought he had discovered the desired theory. It had two unanticipated ingredients. One (*3*) was

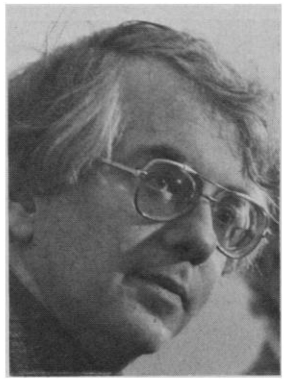
0036-8075/79/1214-1290$00.50/0

another heavy vector meson, the Z. The other (4) was a set of complicated interactions among the photon, W, and Z—interactions that were characteristic of a kind of theory called gauge field theory, invented for other purposes 6 years earlier (5).

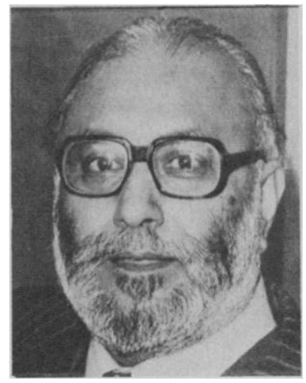
Glashow had reached gauge field theory starting from renormalizability. Four years later, Salam and J. C. Ward, ignorant of Glashow's 1960 work, found the same theory while pursuing their long-held ambition to explain all fundamental interactions as gauge field theory (6). In retrospect, it is not surprising that the same destination should have been reached from two such different starting points, for we now believe that the correct theory is both renormalizable and a gauge theory. However, it is *not* the theory found by Glashow, Salam, and Ward; despite Glashow's high hopes in 1960, that theory is not renormalizable. As we now know, it would be renormalizable, all potential infinities would cancel, if physics were absolutely symmetric among the three vector mesons (photon, W, and Z). In the 1960 theory, the symmetry is spoiled by the different masses assigned the different mesons, and this in turn spoils the cancellation of infinities. (Actually, to make all infinities cancel, the masses not only have to be all equal, they have to be all zero.)

It was not until 1967 that a theory that overcame this difficulty was finally published, by Weinberg (7); shortly afterwards, Salam announced his independent discovery of the same theory (8). The earlier careers of both Weinberg and Salam could be read as preparation for this work. From the early 1950's, Salam had been a major figure in both weak-interaction theory and renormalization theory. Weinberg's 1957 thesis was on renormalization and weak interactions, and he had gone on to make important contributions to both fields. Most important, though, both had been active in the development of the theory of hidden symmetry; it was here that the solution was found (9).

Hidden symmetry is best explained by a parable. The parable involves a man who lives inside a gigantic ferromagnet—a very large array of magnetic dipoles, interacting in such a way that nearby dipoles tend to align. Note that this description is completely symmetric with respect to spatial direction; there is nothing in it to distinguish north from east. The ground state of the ferromagnet, its state of lowest energy, is one in which all the dipoles are aligned in some direction. To be definite, let us assume it is north. Now a man living in the ferromagnet will



Sheldon Glashow



Abdus Salam

Steven Weinberg

not observe the directional symmetry of the laws of nature; for him, north will be a very special direction, the direction in which there is a huge magnetic field that critically affects all of his experiments. It does no good to suggest to him that it would be just as consistent with the laws of nature for the magnetic field to point east; it is a very large ferromagnet, and it is energetically impossible for him to attempt to realign the magnet. Only by a long and exacting sequence of experiment and analysis will he gain an understanding of the microscopic structure of the world in which he lives and discover the directional symmetry that is hidden from him.

The reading of the parable is this: The directional symmetry is the symmetry between the vector mesons, the magnet is the universe, and the man in the magnet is you and I. The ground state of the universe, what we think of as empty space and what high-energy theorists call the vacuum state, has a structure as complex as the ground state of the magnet. Like the ground state of the magnet, its structure is determined by the form of the laws of nature, and it can hide a symmetry those very laws possess. This is how the Weinberg-Salam theory attains renormalizability. The cancellation of potential infinities is not spoiled because the symmetry between the vector mesons is not spoiled. It is only hidden.

This was Weinberg and Salam's masterstroke. Hidden symmetry had given them not only a renormalizable theory, but also a truly unified one. Its predecessors had been only partially unified; the weak and electromagnetic interactions were similar, but different; there was something in the laws of nature that distinguished the photon from the other vector mesons. In the Weinberg-Salam theory this is not so; the laws of nature no more know the photon as a special particle than the laws of the ferromagnet know north as a special direction.

There were only two things wrong with the Weinberg-Salam theory. (i) In its original form, the theory treated leptons only. (Indeed, the title of Weinberg's paper was "A theory of leptons.") It was not clear how to extend the theory to strongly interacting particles. (ii) Nobody paid any attention to it. Rarely has so great an accomplishment been so widely ignored. Here is a census of citations of Weinberg's 1967 paper as recorded in 7 years of *Science Citation Index*: 1967, 0; 1968, 0; 1969, 0; 1970, 1; 1971, 4; 1972, 64; and 1973, 162. These include citations by Weinberg himself (*10*).

Weinberg was of two minds about his theory. Friends recall that when it was published he told them it was the best thing he had ever done. At the same time, though, he worried whether the same mechanisms that hid the symmetry of the theory could also spoil its renormalizability. He struggled intermittently with the problem for years, but never obtained any results he felt were worth publishing. He spent most of his time working on other things (*11*).

Likewise for Salam. When he presented his theory, he expressed his belief in its renormalizability, and even gave the kernel of a correct argument for his belief (*12*). Nevertheless, 2 years later, he spoke "On renormalization constants and inter-relation of fundamental forces" at an international conference (*13*). He did not mention his theory of the weak and electromagnetic interactions; he was working on other things.

It is clear from the citation census that something happened in late 1971. What happened was that Gerard 't Hooft wrote a paper on the renormalization of gauge theories that revealed Weinberg and Salam's frog to be an enchanted prince (*14*). Although 't Hooft's work fell short of a complete proof of renormalizability, it went far enough to transform the subject. Before 't Hooft, the renormalizability of the theory was conjectural; after 't Hooft, it was the nonrenormalizability that was conjectural (*15*).

The 1971 revival of the theory found

the extension to strongly interacting particles ready to hand. In 1970, Glashow, John Iliopoulos, and Luciano Maiani had performed an extraordinary analysis (*16*). They did not know how to build a renormalizable weak-interaction theory; the Weinberg-Salam theory was then sunk in deepest obscurity. Nevertheless, they knew that something must render higher-order weak-interaction effects finite, and they argued that whatever that something was, it would run into experimental contradictions when extended to the strongly interacting particles, unless the theory of the strong interaction itself obeyed certain constraints. In particular, there must exist kinds of strongly interacting particles that had not yet been observed, which they called charmed particles (*17*).

The indirection of the argument deserves emphasis. New kinds of strongly interacting particles were predicted, not on the basis of strong-interaction theory per se, but in order to make strong-interaction theory consistent with a weak-interaction theory that (so far as Glashow *et al.* knew) had not yet been invented. Despite its indirection, Glashow found the argument absolutely convincing; he was unwavering in his belief that charmed particles must exist and that experimenters must look for them. Of course, as any newspaper reader knows, they do exist, but I will forgo telling the story of their discovery here (*18*).

By 1971, then, all the pieces had been assembled. Weinberg-Salam plus 't Hooft plus Glashow-Iliopoulos-Maiani had finally yielded the theory striven for since 1958. Experimental techniques had also advanced enormously; there existed neutrino beams, electron accelerators, detection apparatus vastly more powerful than anything available in the 1950's. Weinberg wrote that we now had a structure of great theoretical appeal "neither confirmed nor refuted by present data." It predicted novel effects, "effects which are just on the verge of observability" (*19*). The ball was in the experimenters' court.

In 1973, experiments at CERN, near Geneva, and Fermilab, near Chicago, detected neutral-current events (that is, weak interactions involving exchange of the Z vector meson) of a form and magnitude consistent with the theory. The next 5 years were a confusing period of exhilaration and disappointment, alarms and excursions. Experiment confirmed the theory; experiment denied the theory. Enormous theoretical effort was devoted to producing grotesque mutant versions of the theory consistent with the new experimental results; the new experiments were shown to be in error; the mutants were slain (*20*). In the last few years, though, the experimental situation seems to have stabilized in agreement with the original 1971 version of the theory (*21*). The Weinberg-Salam model is now the standard theory of the electroweak interaction.

Even before it was confirmed experimentally, the standard theory had effects outside the domain of the weak and electromagnetic interactions. It inspired a surge of activity in gauge theory in general; in 1973, this led to the discovery that certain remarkable phenomena observed in electron-nucleon scattering could be explained by assuming that the strong interaction itself was also a gauge theory. From this observation have grown many (though not all) of our reasons for believing in the current leading candidate for the theory of the strong interaction quantum chromodynamics, the theory of quarks and gluons.

If the weak, electromagnetic, and strong interactions are all three gauge theories, then it is tempting to envision unifying them all, in the same way the standard theory unifies electromagnetic and weak interactions. (The parallelism is not precise; in this case, renormalizability does not force unification.) Glashow, Salam, and Weinberg have all been leaders in the development and investigation of such grand unified theories. Unfortunately, both decisive experimental tests and stringent theoretical constraints are hard to find here. Thus, although grand unification in general is an extraordinarily attractive idea, the particulars of how it happens remain mysterious.

Only gravity has resisted the armies of unification. Despite strenuous assaults, it remains in adamant isolation: not a gauge theory (at least not in the same way other interactions are gauge theories) and not renormalizable. I suspect the understanding of gravity will require some profoundly new physics, physics that will be in its way as original, ingenious, beautiful, and deep as the physics which won this year's Nobel prize.

SIDNEY COLEMAN
Stanford Linear Accelerator Center, Stanford, California 94305

References and Notes

1. "It is of little value to have a potentially renormalizable theory of beta processes without the possibility of a renormalizable electrodynamics. We should care to suggest that a fully acceptable theory of these interactions may only be achieved if they are treated together." [S. L. Glashow, thesis, Harvard University (1958), p. 75].
2. J. Schwinger, *Ann. Phys.* **2**, 407 (1957).
3. S. L. Glashow, *Nucl. Phys.* **22**, 579 (1961). This paper was received by the journal in September 1960, the date I have used above.
4. S. L. Glashow, thesis (*1*); *Nucl. Phys.* **10**, 107 (1959).
5. Gauge field theories were invented by C. N. Yang and R. Mills [*Phys. Rev.* **96**, 190 (1954)]. The suggestion that the weak interaction might be a gauge field theory had been made earlier by S. Bludman [*Nuovo Cimento* **9**, 433 (1958)], but Bludman's theory did not include electromagnetism.
6. A. Salam and J. C. Ward, *Phys. Lett.* **13**, 168 (1964). Salam and Ward had made a very early attempt to construct a unified gauge theory of the weak and electromagnetic interactions [*Nuovo Cimento* **11**, 568 (1959)], and, long before the developments discussed at the end of this article, they had proposed a gauge theory of the strong interaction [*ibid.* **20**, 419 (1961)]. The difference between Glashow's viewpoint and Salam and Ward's is striking. Glashow knew of the connection with gauge theories, but treated it as a throwaway line, remarking in his 1959 paper that what he found was "formally identical to the vector meson interaction observed by Yang and Mills in another context." Complementarily, Salam and Ward mentioned renormalizability only in a throwaway line in their 1959 paper, remarking of gauge theory that "it is perhaps the only theory of charged vector mesons that can be renormalized."
7. S. Weinberg, *Phys. Rev. Lett.* **19**, 1264 (1967).
8. A. Salam, in *Elementary Particle Theory, Proceedings of the Eighth Nobel Symposium*, N. Svartholm, Ed. (Wiley-Interscience, New York, 1968).
9. What I call hidden symmetry is more usually called spontaneous symmetry breakdown. The idea was introduced into high-energy theory by J. Goldstone, Y. Nambu, and G. Jona-Lasinio. The critical extension to gauge theories was done by F. Englert and R. Brout, P. Higgs, and G. Guralnik, C. Hagen, and T. Kibble. Weinberg and Salam collaborated with Goldstone on one of the classic papers of the field [*Phys. Rev.* **127**, 965 (1962)]. In 1967, Weinberg was fresh from his triumphant sequence of applications of hidden symmetry to low-energy meson phenomena. The last paper of this set [*Phys. Rev. Lett.* **18**, 507(1967)] is much concerned with the interplay between vector mesons and hidden symmetries; it clearly foreshadows the great work on the weak and electromagnetic interactions. Weinberg discusses these connections in his review pa[illegible] [*Rev. Mod. Phys.* **46**, 255 (1974)].
10. In contrast, in 1968 there were 134 citations of Weinberg's 1967 paper mentioned in (*9*).
11. Among them was the ABM controversy. After 't Hooft had transformed the subject, Weinberg reported the unpublished results referred to above [*Phys. Rev. Lett.* **27**, 1688 (1971)].
12. See especially his response to E. C. G. Sudarshan's questions at the Nobel Symposium (*9*).
13. Seminar on Fundamental Problems of Elementary Particle Theory, Kiev, 1970.
14. G. 't Hooft, *Nucl. Phys. B* **35**, 167 (1971). 't Hooft at this time was a graduate student at Utrecht working under the direction of Tini Veltman. A month or so before he announced his result, Veltman and I met at Marseilles. Veltman said, "I have a student who has constructed a renormalizable theory of charged vector mesons." I said, "I don't believe it."
15. Renormalizability was later proved rigorously by B. W. Lee and J. Zinn-Justin and by 't Hooft and Veltman.
16. S. Glashow, J. Iliopoulos, L. Maiani, *Phys. Rev. D* **2**, 1285 (1970).
17. Glashow had actually raised the possibility of charmed particles much earlier, in a paper written with J. Bjorken [*Phys. Lett.* **11**, 255 (1964)]. In 1964 charm was just a possibility, though; in 1970 it was a necessity.
18. Part of this story is told in J. Bjorken [*Science* **194**, 826 (1976)].
19. Both quotations are from S. Weinberg [*Phys. Rev. D* **5**, 1412 (1972)]. Even at this late date a consensus had not quite formed. For example, in this paper Weinberg is very serious about the Weinberg-Salam theory, but still not sure that charm is the solution for the strongly interacting particles. At roughly the same time, Glashow and collaborators were taking the complementary position—strong on charm, uncommitted on Weinberg-Salam.
20. This happened at least three times: with high-y anomalies, with trimuons, and with atomic parity violation.
21. Strictly speaking, the current theory is a generalization of the 1971 theory; it contains additional entities, generalizations of charmed particles. Such an extension is a minimal modification of the theory, and I do not think anyone would be particularly disturbed if new experiments required further extensions of the same sort.