

Nobel Lecture: The discovery of asymptotic freedom and the emergence of QCD*

David J. Gross

Kavli Institute for Theoretical Physics, UCSB, Santa Barbara, California 93106, USA

(Published 7 September 2005)

INTRODUCTION

The progress of science is much more muddled than is depicted in most history books. This is especially true of theoretical physics, partly because history is written by the victorious. Consequently, historians of science often ignore the many alternate paths that people wandered down, the many false clues they followed, the many misconceptions they had. These alternate points of view are less clearly developed than the final theories, harder to understand, and easier to forget, especially as these are viewed years later, when it all really does make sense. Thus reading history one rarely gets the feeling of the true nature of scientific development, in which the element of farce is as great as the element of triumph.

The emergence of QCD is a wonderful example of the evolution from farce to triumph. During a very short period, a transition occurred from experimental discovery and theoretical confusion to theoretical triumph and experimental confirmation. In this Nobel lecture I shall describe the turn of events that led to the discovery of asymptotic freedom, which in turn led to the formulation of QCD, the final element of the remarkably comprehensive theory of elementary particle physics—the Standard Model. I shall then briefly describe the experimental tests of the theory and the implications of asymptotic freedom.

PARTICLE PHYSICS IN THE 1960s

The early 1960s, when I started my graduate studies at UC Berkeley, were a period of experimental supremacy and theoretical impotence. The construction and utilization of major accelerators were proceeding at full steam. Experimental discoveries and surprises appeared every few months. There was hardly any theory to speak of. The emphasis was on phenomenology, and there were only small islands of theoretical advances here and there. Field theory was in disgrace; *S*-matrix theory was in full bloom. Symmetries were all the rage. Of the four forces observed in nature, only gravity and electromagnetism were well understood. The other two forces, the weak force responsible for radioactivity and the strong nuclear force that operated within the nucleus, were

largely mysterious. Particle physics was divided into the study of the weak and the strong interactions, the two mysterious forces that operate within the nucleus. In the case of the weak interactions, there was a rather successful phenomenological theory, but not much new data. The strong interactions were where the experimental and theoretical action was, particularly at Berkeley. They were regarded as especially unfathomable. In hindsight this was not surprising since nature was hiding her secrets. The basic constituents of hadrons (strongly interacting particles) were invisible. We now know that these are quarks, but no one had ever seen a quark, no matter how hard protons were smashed into protons. Furthermore, the “color” charges we now know are the source of the chromodynamic fields, the analogs of the electric charge, were equally invisible. The prevalent feeling was that it would take a very long time to understand the nuclear force and that it would require revolutionary concepts. Freeman Dyson had asserted that “the correct theory will not be found in the next hundred years.” For a young graduate student, such as myself, this was clearly the biggest challenge.

QUANTUM FIELD THEORY

Quantum field theory was originally developed for the treatment of electrodynamics, immediately after the completion of quantum mechanics and the discovery of the Dirac equation. It seemed to be the natural tool for describing the dynamics of elementary particles. The application of quantum field theory to the nuclear forces had important early success. Fermi formulated a powerful and accurate phenomenological theory of beta decay, which (though deficient at high energy) was to serve as a framework for exploring the weak interactions for three decades. Yukawa proposed a field theory to describe the nuclear force and predicted the existence of heavy mesons, which were soon discovered. On the other hand, quantum field theory was confronted from the beginning with severe difficulties. These included the infinities that appeared as soon as one went beyond lowest-order perturbation theory, as well as the lack of any nonperturbative tools. By the 1950s the suspicion of field theory had deepened to the point that a powerful dogma emerged—that field theory was fundamentally wrong, especially in its application to the strong interactions.

The renormalization procedure, developed by R. Feynman, J. Schwinger, S. Tomonaga, and F. Dyson, which had eliminated the ubiquitous infinities that oc-

*The 2004 Nobel Prize in Physics was shared by David J. Gross, H. David Politzer, and Frank Wilczek. This lecture is the text of Professor Gross's address in conjunction with the award.

curred in calculations by expressing physical observables in terms of physical parameters, was spectacularly successful in quantum electrodynamics. However, the physical meaning of renormalization was not truly understood. The feeling of most was that renormalization was a trick. This was especially the case for the pioneering inventors of quantum field theory. They were prepared at the first appearance of divergences to renounce their belief in quantum field theory and to brace for the next revolution. However, it was also the feeling of the younger leaders of the field, who had laid the foundations of perturbative quantum field theory and renormalization in the late 1940s. The prevalent feeling was that renormalization simply swept the infinities under the rug, but that they were still there and rendered the notion of local fields meaningless. To quote Feynman, speaking at the 1961 Solvay conference (Feynman, 1962), “I still hold to this belief and do not subscribe to the philosophy of renormalization.”

Field theory was almost totally perturbative at that time; all nonperturbative techniques that had been tried in the 1950s had failed. The path integral, developed by Feynman in the late 1940s, which later proved so valuable for a nonperturbative formulation of quantum field theory as well as a tool for semiclassical expansions and numerical approximations, was almost completely forgotten. In a sense, the Feynman rules were too successful. They were an immensely useful, picturesque, and intuitive way of performing perturbation theory. However, these alluring qualities also convinced many that all that was needed from field theory were these rules. They diverted attention from the nonperturbative dynamical issues facing field theory. In my first course on quantum field theory at Berkeley in 1965, I was taught that Field Theory = Feynman Rules. Today we know that there are many phenomena, especially confinement in QCD, that cannot be understood perturbatively.

In the United States, the main reason for the abandonment of field theory for the strong interactions was simply that one could not calculate. American physicists are inveterate pragmatists. Quantum field theory had not proved to be a useful tool with which to make contact with the explosion of experimental discoveries. The early attempts in the 1950s to construct field theories of the strong interactions were total failures. In hindsight this was not surprising since a field theory of the strong interactions faced two enormous problems. First, which fields to use? Following Yukawa, the first attempts employed pion and nucleon fields. Soon, with the rapid proliferation of particles, it became evident that nothing was special about the nucleon or the pion. All the hadrons, the strange baryons and mesons as well as the higher-spin recurrences of these, appeared to be equally fundamental. The obvious conclusion that all hadrons were composites of more fundamental constituents was thwarted by the fact that, no matter how hard hadrons were smashed into one another, one had not been able to liberate these hypothetical constituents. This was not analogous to the paradigm of atoms made of nucleons and electrons or of nuclei composed of nucleons. The

idea of permanently bound, confined constituents was unimaginable at the time. Second, since the pion-nucleon coupling was so large, perturbative expansions were useless, and all attempts at nonperturbative analysis were unsuccessful.

In the case of the weak interactions, the situation was somewhat better. Here one had an adequate effective theory, the four-fermion Fermi interaction, which could be usefully employed, using perturbation theory to lowest order, to organize and understand the emerging experimental picture of the weak interactions. The fact that this theory was nonrenormalizable meant that beyond the Born approximation it lost all predictive value. This disease increased the suspicion of field theory. Yang-Mills theory, which had appeared in the mid-1950s, was not taken seriously. Attempts to apply Yang-Mills theory to the strong interactions focused on elevating global flavor symmetries to local gauge symmetries. This was problematic since these symmetries were not exact. In addition non-Abelian gauge theories apparently required massless vector mesons—clearly not a feature of the strong interactions.

In the Soviet Union field theory was under even heavier attack, for somewhat different reasons. Landau and collaborators, in the late 1950s, studied the high-energy behavior of quantum electrodynamics. They explored the relation between the physical electric charge and the bare electric charge as seen at infinitesimally small distances. The fact that the electric charge in QED depends on the distance at which we measure it is due to “vacuum polarization.” The vacuum, the ground state of a relativistic quantum-mechanical system, should be thought of as a medium consisting of virtual particles. In QED the vacuum contains virtual electron-positron pairs. If a charge is inserted into this dielectric medium, it distorts, or polarizes the virtual dipoles and this will screen the charge. Consequently the charge seen at some distance will be reduced in magnitude, and the farther one goes the smaller the charge. We can introduce the notion of an effective charge, $e(r)$, which determines the force at a distance r . As r increases, there is more screening medium, thus $e(r)$ decreases with increasing r , and correspondingly increases with decreasing r . The beta function, which is minus the logarithmic derivative of the charge with respect to distance, is thus positive. Landau and colleagues concluded, on the basis of their approximations, that this effect is so strong that the physical charge, as measured at any finite distance, would vanish for any value of the bare charge. They stated: “We reach the conclusion that within the limits of formal electrodynamics a point interaction is equivalent, for any intensity whatever, to no interaction at all” (Landau and Pomeranchuk, 1955).

This is the famous problem of zero charge, a startling result that implied for Landau that “weak-coupling electrodynamics is a theory, which is, fundamentally, logically incomplete” (Landau, 1955). This problem occurs in any non-asymptotically-free theory. Even today, many of us believe that many non-asymptotically-free theories, such as QED, are inconsistent at very high energies.

In the case of QED this is only an academic problem, since the trouble shows up only at enormously high energy. However, Landau believed that this phenomenon was more general and would occur in all field theories. Why? First, every theory they looked at had this property. But more importantly, I think, dielectric screening is a natural physical explanation of charge renormalization, and they were unaware of any simple physical reason for the opposite effect. Thus they assumed that the problem of zero charge would arise in any field theory of the strong interaction, but here it was an immediate catastrophe. In the Soviet Union this was thought to be a compelling reason why field theory was wrong, and certainly inappropriate for the strong force. Landau decreed that “We are driven to the conclusion that the Hamiltonian method for strong interaction is dead and must be buried, although of course with deserved honor” (Landau, 1960).

Under the influence of Landau and Pomeranchuk, a generation of physicists was forbidden to work on field theory. Why did the discovery of the zero-charge problem not inspire a search for asymptotically free theories that would be free of this disease? The answer, I think, is twofold. First, many other theories were explored—in each case they behaved as QED. Second, Landau had concluded that this problem was inherent in any quantum field theory, that an asymptotically free theory could not exist. V. S. Vanyashin and M. V. Teren'tev carried out a calculation of the charge renormalization of charged vector mesons in 1964 (Teren'tev and Vanyashin, 1965). They got the magnitude wrong but did get the correct sign and concluded that the result was absurd. They attributed this wrong sign to the nonrenormalizability of charged-vector-meson theory.

THE BOOTSTRAP

If field theory could not provide the theoretical framework for the strong interactions what could? In the early sixties a radically different approach emerged— S -matrix theory and the bootstrap. The bootstrap theory rested on two principles, both more philosophical than scientific. First, local fields were not directly measurable. Thus they were unphysical and meaningless. Instead, one should formulate the theory using the observable S -matrix elements measured in scattering experiments. Microscopic dynamics was renounced. Field theory was to be replaced by S -matrix theory; a theory based on general principles, such as unitarity and analyticity, but with no fundamental microscopic Hamiltonian. The basic dynamical idea was that there was a unique S matrix that obeyed these principles. It could be determined without requiring any fundamental constituents or equations of motion (Chew, 1963). In hindsight, it is clear that the bootstrap was born from the frustration of being unable to calculate anything using field theory. All models and approximations produced conflict with some dearly held principle. If it was so difficult to construct an S matrix that was consistent with sacred principles then maybe these principles had a unique manifestation. The

second principle of the bootstrap was that there were no elementary particles. The way to deal with the increasing number of candidates for elementary status was to proclaim that all were equally fundamental; all were dynamical bound states of each other. This was called Nuclear Democracy and was a response to the proliferation of candidates for fundamental building blocks.

S -matrix theory had some notable successes, such as dispersion relations and the development of Regge pole theory. However, there were drawbacks to a theory that was based on the principle that there was no theory, at least in the traditional sense. Nonetheless, until 1973 it was not thought proper to use field theory without apologies. For example, as late as the National Accelerator Laboratory conference of 1972, Murray Gell-Mann ended his talk on quarks with the summary: “Let us end by emphasizing our main point, that it may well be possible to construct an explicit theory of hadrons, based on quarks and some kind of glue, treated as fictitious, but with enough physical properties abstracted and applied to real hadrons to constitute a complete theory. Since the entities we start with are fictitious, there is no need for any conflict with the bootstrap or conventional dual parton point of view” (Gell-Mann, 1973).

SYMMETRIES

If dynamics was forbidden, one could at least explore the symmetries of the strong interactions. The biggest advance of the early 1960s was the discovery of an approximate symmetry of hadrons, $SU(3)$, by Gell-Mann and Neeman (1964) and then the beginning of the understanding of spontaneously broken chiral symmetry. Since the relevant degrees of freedom, especially color, were totally hidden from view due to confinement, the emphasis was on flavor, which was directly observable. This emphasis was enhanced because of the success of $SU(3)$. Nowadays we realize that $SU(3)$ is an accidental symmetry, which arises simply because a few quarks (the up, down, and strange quarks) are relatively light compared to the scale of the strong interactions. At the time it was regarded as a deep symmetry of the strong interactions, and many attempts were made to generalize it and use it as a springboard for a theory of hadrons. The most successful attempt was Gell-Mann's algebra of currents, a program for abstracting relations from a field theory, keeping the ones that might be generally true and then throwing the field theory away, “In order to obtain such relations that we conjecture to be true, we use the method of abstraction from a Lagrangian field-theory model. In other words, we construct a mathematical theory of the strongly interacting particles, which may or may not have anything to do with reality, find suitable algebraic relations that hold in the model, postulate their validity, and then throw away the model. We may compare this process to a method sometimes employed in French cuisine: a piece of pheasant meat is cooked between two slices of veal, which are then discarded” (Gell-Mann, 1964a). This paper made quite an impres-

sion, especially on impoverished graduate students like me, who could only dream of eating such a meal. It was a marvelous approach. It gave one the freedom to play with the forbidden fruit of field theory and abstract what one wanted from it, all without having to believe in the theory. The only problem was that it was not clear what principle determined what to abstract.

The other problem with this approach was that it diverted attention from dynamical issues. The most dramatic example of this is Gell-Mann and George Zweig's hypothesis of quarks, the most important consequence of the discovery of $SU(3)$ (Gell-Mann, 1964b; Zweig, 1964). The fact was that hadrons looked as if they were composed of (colored) quarks whose masses (either the current quark masses or the constituent quark masses) were quite small. Color had been introduced by O. W. Greenberg (1964), Y. Nambu (1965, 1968) and M. Y. Han and Nambu (1965). Nambu's motivation for color was twofold; first to offer an explanation of why only (what we would now call) color singlet hadrons exist by postulating a strong force (but with no specification as to what kind of force) coupled to color, which was responsible for the fact that color-neutral states were lighter than colored states. The second motivation, explored with Han, was the desire to construct models in which the quarks had integer-valued electric charges. Greenberg's motivation was to explain the strange statistics of non-relativistic quark-model hadronic bound states (a concern of Nambu's as well). He introduced parastatistics for this purpose, which solved the statistics problem, but clouded the dynamical significance of this quantum number.

Yet quarks had not been seen, even when energies were achieved that were ten times the threshold for their production. The nonrelativistic quark model simply did not make sense. The conclusion was that quarks were fictitious, mathematical devices. If one had believed in an underlying field theory it would be hard to maintain this attitude, but it was certainly consistent with the bootstrap. With this attitude one could ignore the apparently insoluble dynamical problems that arose if one tried to imagine that quarks were real. This attitude towards quarks persisted until 1973 and beyond. Quarks clearly did not exist as real particles; therefore they were fictitious devices (see Gell-Mann above). One might "abstract" properties of quarks from some model, but one was not allowed to believe in their reality or to take the models too seriously. For many this smelled fishy. I remember very well Steve Weinberg's reaction to the sum rules Curtis Callan and I had derived using the quark-gluon model. I described my work on deep-inelastic scattering sum rules to Weinberg at a Junior Fellows dinner at Harvard. I explained how the small longitudinal cross section observed at SLAC could be interpreted, on the basis of our sum rule, as evidence for quarks. Weinberg was emphatic that this was of no interest since he did not believe anything about quarks.

MY ROAD TO ASYMPTOTIC FREEDOM

I was a graduate student at Berkeley at the height of the bootstrap and S -matrix theory. My Ph.D. thesis was written under the supervision of Geoff Chew, the main guru of the bootstrap, on multibody N/D equations. I can remember the precise moment at which I was disillusioned with the bootstrap program. This was at the 1966 Rochester meeting, held at Berkeley. Francis Low, in the session following his talk, remarked that the bootstrap was less of a theory than a tautology, "I believe that when you find that the particles that are there in S -matrix theory, with crossing matrices and all the formalism, satisfy all these conditions, all you are doing is showing that the S matrix is consistent with the world the way it is; that is, the particles have put themselves there in such a way that it works out, but you have not necessarily explained that they are there" (Low, 1967). For example, the then popular finite-energy sum rules (whereby one derived relations for measurable quantities by saturating dispersion relations with a finite number of resonance poles on the one hand and relating these to the assumed Regge asymptotic behavior on the other) were not so much predictive equations as merely checks of axioms (analyticity, unitarity) using models and fits of experimental data. I was very impressed with this remark and longed to find a more powerful dynamical scheme. This was the heyday of current algebra, and the air was buzzing with marvelous results. I was very impressed by the fact that one could assume a certain structure of current commutators and derive measurable results. The most dramatic of these was the Adler-Weisberger relation that had just appeared (Adler, 1965; Weisberger, 1966). Clearly the properties of these currents placed strong restrictions on hadronic dynamics. The most popular scheme then was current algebra. Gell-Mann and R. Dashen were trying to use the commutators of certain components of the currents as a basis for strong-interaction dynamics (Dashen and Gell-Mann, 1966). After a while I concluded that this approach was also tautological. All it did was test the validity of the symmetries of the strong interactions. This was apparent for vector $SU(3)$, but was also true of chiral $SU(3)$, especially as Weinberg and others interpreted the current-algebra sum rules as low-energy theorems for Goldstone bosons. This scheme could not be a basis for a complete dynamical theory. I therefore studied the less understood properties of the algebra of local current densities. These were model dependent, but that was fine. They therefore might contain dynamical information that went beyond statements of global symmetry. Furthermore, as was soon realized, one could check one's assumptions about the structure of local current algebra by deriving sum rules that could be tested in deep-inelastic lepton-hadron scattering experiments. J. Bjorken's 1966 paper, on the application of $U(6) \times U(6)$, particularly influenced me (Bjorken, 1966, 1968). In the spring of 1968 Curtis Callan and I proposed a sum rule to test the then popular "Sugawara model," a dynamical model of local currents, in which the energy-momentum

tensor was expressed as a product of currents. The hope was that the algebraic properties of the currents and the expression for the Hamiltonian in terms of these would be enough to have a complete theory. Our goal was slightly more modest—to test the hypothesis by exploiting the fact that in this theory the operator-product expansion of the currents contained the energy-momentum tensor with a known coefficient. Thus we could derive a sum rule for the structure functions that could be measured in deep-inelastic electron-proton scattering (Callan and Gross, 1968a). In the fall of 1968, Bjorken noted that this sum rule, as well as dimensional arguments, would suggest the scaling of deep-inelastic scattering cross sections. This prediction was shortly confirmed by the new experiments at SLAC, which were to play such an important role in elucidating the structure of hadrons (Bloom *et al.*, 1969). Shortly thereafter Callan and I discovered that by measuring the ratio $R = \sigma_L / \sigma_T$ [where σ_L (σ_T) is the cross section for the scattering of longitudinal (transverse) polarized virtual photons], one could determine the spin of the charged constituents of the nucleon. We evaluated the moments of the deep-inelastic structure functions in terms of the equal-time commutators of the electromagnetic current using specific models for these—the algebra of fields in which the current was proportional to a spin-one field on the one hand, and the quark-gluon model on the other. In this popular model quarks interacted through an Abelian gauge field (which could, of course, be massive) coupled to baryon number. The gauge dynamics of the gluon had never been explored, and I do not think that the model had been used to calculate anything until then. We discovered that R depended crucially on the spin of the constituents. If the constituents had spin zero or one, then $\sigma_T = 0$, but if they had spin 1/2, then $\sigma_L = 0$ (Callan and Gross, 1968b). This was a rather dramatic result. The experiments quickly showed that σ_L was very small.

These SLAC deep-inelastic scattering experiments had a profound impact on me. They clearly showed that the proton behaved, when observed over short times, as if it was made out of pointlike objects of spin one-half. In the spring of 1969, which I spent at CERN, C. Llewellyn-Smith and I analyzed the sum rules that followed for deep-inelastic neutrino-nucleon scattering using similar methods (Gross and Llewellyn-Smith, 1969). We were clearly motivated by the experiments that were then being performed at CERN. We derived a sum rule that measured the baryon number of the charged constituents of the proton. The experiments soon indicated that the constituents of the proton had baryon number 1/3, in other words again they looked like quarks. I was then totally convinced of the reality of quarks. They had to be more than just mnemonic devices for summarizing hadronic symmetries, as they were then universally regarded. They had to be physical pointlike constituents of the nucleon. But how could that be? Surely strong interactions must exist between the quarks that would smear out their pointlike behavior.

After the experiments at SLAC, Feynman came up

with his parton picture of deep-inelastic scattering (Feynman, 1969), a very picturesque and intuitive way of describing deep-inelastic scattering in terms of assumed pointlike constituents—partons. It complemented the approach to deep-inelastic scattering based on the operator product of currents, and had the advantage of being extendible to other processes (Drell and Yan, 1971). The parton model allowed one to make predictions with ease, ignoring the dynamical issues at hand. I felt more comfortable with the approach based on assuming properties of current products at short distances, and felt somewhat uneasy about the extensions of the parton model to processes that were not truly dominated by short-distance singularities. At CERN I studied, with Julius Wess, the consequences of exact scale and conformal invariance (Gross and Wess, 1970). However, I soon realized that in a field-theoretic context only a free, noninteracting theory could produce exact scaling. This became very clear to me in 1969, when I came to Princeton, where my colleague C. Callan (and K. Symansik) had rediscovered the renormalization-group equations, which they presented as a consequence of a scale invariance anomaly (Callan, 1970; Symansik, 1970). Their work made it abundantly clear that once one introduced interactions into the theory, scaling, as well as my beloved sum rules, went down the tube. Yet the experiments indicated that scaling was in fine shape. One could hardly turn off the interactions between the quarks, or make them very weak, since then one would expect hadrons to break up easily into their quark constituents, and no one ever observed free quarks. This paradox and the search for an explanation of scaling were to preoccupy me for the following four years.

HOW TO EXPLAIN SCALING

About the same time that all this was happening, string theory was discovered, in one of the most bizarre turns of events in the history of physics. In 1968 G. Veneziano (1968) came up with a remarkably simple formula that summarized many features of hadronic scattering, with Regge asymptotic behavior in one channel and narrow resonance saturation in the other. This formula was soon generalized to multiparticle S -matrix amplitudes and attracted much attention. The dual-resonance model was born, the last serious attempt to implement the bootstrap. It was only truly understood as a theory of quantized strings in 1972. I worked on this theory for two years, at CERN and then at Princeton with Schwarz and Neveu. At first I felt that this model, which captured many of the features of hadronic scattering, might provide the long sought alternative to a field theory of the strong interactions. However, by 1971 I realized that there was no way that this model could explain scaling, and I felt strongly that scaling was the paramount feature of the strong interactions. In fact the dual-resonance model led to incredibly soft behavior at large momentum transfer, quite the opposite of the hard scaling observed. Also, it required for consistency features that were totally unrealistic for the strong interactions—

massless vector and tensor particles. These features later became the motivation for the hope that string theory might provide a comprehensive and unified theory of all the forces of nature. This hope remains strong. However, the relevant energy scale is not 1 GeV but rather 10^{19} GeV!

The data on deep-inelastic scattering were getting better. No violations of scaling were observed, and the free-field-theory sum rules worked. I remember well the 1970 Kiev conference on high-energy physics. There I met S. Polyakov and S. Migdal, uninvited but already impressive participants at the meeting. Polyakov, Migdal, and I had long discussions about deep-inelastic scattering. Polyakov knew all about the renormalization group and explained to me that naive scaling could be right. Because of renormalization the dimensions of operators change with the scale of the physics being probed. Not only that, dimensionless couplings also change with scale. They approach at small distances fixed-point values that are generically those of a strongly coupled theory, resulting in large anomalous scaling behavior quite different from free-field-theory behavior. I retorted that the experiments showed otherwise. He responded that this behavior contradicts field theory. We departed, he convinced, as many were, that experiments at higher energies would change, I that the theory would have to be changed. The view that the scaling observed at SLAC was not a truly asymptotic phenomenon was rather widespread. The fact that scaling set in at rather low momentum transfers, “precocious scaling,” reinforced this view. Thus the cognoscenti of the renormalization group (Wilson, Polyakov, and others) believed that the noncanonical scaling indicative of a nontrivial fixed point of the renormalization group would appear at higher energies.

Much happened during the next two years. Gerard 't Hooft's spectacular work on the renormalizability of Yang-Mills theory reintroduced non-Abelian gauge theories to the community ('t Hooft, 1967). The electroweak theory of S. Glashow, S. Weinberg, and A. Salam was revived. Field theory became popular again, at least in application to the weak interactions. The path integral reemerged from obscurity. Kenneth Wilson's development of the operator-product expansion (Wilson, 1971) provided a new tool that could be applied to the analysis of deep-inelastic scattering. The Callan-Symanzik equations simplified the renormalization-group analysis, which was then applied to the Wilson expansion (Symanzik, 1971; Callan, 1972; Christ *et al.*, 1972; Callan and Gross, 1973). The operator-product analysis was extended to the light cone, the relevant region for deep-inelastic scattering (Jackiw *et al.*, 1970; Leutwyler and Stern, 1970; Frishman, 1971; Gross, 1971; Christ *et al.*, 1972). Most important from my point of view was the revival of the renormalization group by Wilson (Wilson and Kogut, 1974). The renormalization group stems from the fundamental work of Stueckelberg and Petermann (1953), Gell-Mann and Low (1954), and Bogoliubov and Shirkov (1959). This work was neglected for many years, partly because it seemed to pro-

vide only information about physics for large spacelike momenta, which are of no direct physical interest. Also, before the discovery of asymptotic freedom, the ultraviolet behavior was not calculable using perturbative methods, and there were no others. Thus it appeared that the renormalization group provided a framework in which one could discuss, but not calculate, the asymptotic behavior of amplitudes in a physically uninteresting region.

THE PLAN

By the end of 1972, I had learned enough field theory, especially renormalization-group methods, to tackle the problem of scaling head on. I decided, quite deliberately, to prove that local field theory could not explain the experimental fact of scaling and thus was not an appropriate framework for the description of the strong interactions. Thus deep-inelastic scattering would finally settle the issue as to the validity of quantum field theory. The plan of the attack was twofold. First, I would prove that “ultraviolet stability,” the vanishing of the effective coupling at short distances, later called asymptotic freedom, was necessary to explain scaling. Second, I would show that there existed no asymptotically free field theories. The latter was to be expected. After all the paradigm of quantum field theory, quantum electrodynamics (QED), was infrared stable; the effective charge grew larger at short distances and no one had ever constructed a theory in which the opposite occurred. If the effective coupling were, contrary to QED, to decrease at short distances, one might explain how the strong interactions turn off in this regime and produce scaling. Indeed, one might suspect that this is the only way to get pointlike behavior at short distances. It was well understood, due to Wilson's work and its application to deep-inelastic scattering, that one might expect to get scaling in a quantum field theory at a fixed point of the renormalization group. However, this scaling would not have canonical, free-field-theory behavior. Such behavior would mean that the scaling dimensions of the operators that appear in the product of electromagnetic currents at lightlike distances had canonical, free-field dimensions. This seemed unlikely. I knew that if the fields themselves had canonical dimensions, then for many theories this implied that the theory was trivial, i.e., free. Surely this was also true if the composite operators that dominated the amplitudes for deep-inelastic scattering had canonical dimensions.

By the spring of 1973, Callan and I had completed a proof of this argument, extending an idea of G. Parisi (1973) to all renormalizable field theories, with the exception of non-Abelian gauge theories. The essential idea was to prove that the vanishing anomalous dimensions of the composite operators, at an assumed fixed point of the renormalization group, implied the vanishing anomalous dimensions of the fields. This then implied that the theory was free at this fixed point. The conclusion was that naive scaling could be explained only if the assumed fixed point of the renormalization

group was at the origin of coupling space, i.e., the theory must be asymptotically free (Callan and Gross, 1973). Non-Abelian gauge theories were not included in the argument since both arguments broke down for these theories. The discovery of asymptotic freedom made this omission irrelevant.

The second part of the argument was to show that there were no asymptotically free theories at all. I had set up the formalism to analyze the most general renormalizable field theory of fermions and scalars—again excluding non-Abelian gauge theories. This was not difficult, since to investigate asymptotic freedom it suffices to study the behavior of the β functions in the vicinity of the origin of coupling-constant space, i.e., in lowest-order perturbation theory (one-loop approximation). I almost had a complete proof but was stuck on my inability to prove a necessary inequality. I discussed the issue with Sidney Coleman, who was spending the spring semester in Princeton. He came up with the missing ingredient and added some other crucial points—and we had a proof that no renormalizable field theory that consisted of theories with arbitrary Yukawa, scalar, or Abelian gauge interactions could be asymptotically free (Coleman and Gross, 1973). A. Zee had also been studying this. He too was well aware of the advantages of an asymptotically free theory and was searching for one. He derived, at the same time, a partial result, indicating the lack of asymptotic freedom in theories with $SU(N)$ -invariant Yukawa couplings (Zee, 1973a).

THE DISCOVERY OF ASYMPTOTIC FREEDOM

Frank Wilczek started work with me in the fall of 1972. He had come to Princeton as a mathematics student, but soon discovered that he was really interested in particle physics. He switched to the physics department, after taking my field theory course in 1971, and started to work with me. My way of dealing with students, then and now, was to involve them closely with my current work and very often to work with them directly. This was certainly the case with Frank, who functioned more as a collaborator than a student from the beginning. I told him about my program to determine whether quantum field theory could account for scaling. We decided that we would calculate the β function for Yang-Mills theory. This was the one hole in the line of argument I was pursuing. It had not been filled largely because Yang-Mills theory still seemed strange and difficult. Few calculations beyond the Born approximation had ever been done. Frank was interested in this calculation for other reasons as well. Yang-Mills theory was already in use for the electroweak interactions, and he was interested in understanding how these behaved at high energy.

Coleman, who was visiting in Princeton, asked me at one point whether anyone had ever calculated the β function for Yang-Mills theory. I told him that we were working on this. He expressed interest because he had asked his student, H. David Politzer, to generalize the mechanism he had explored with Eric Weinberg, that of

dynamical symmetry breaking of an Abelian gauge theory, to the non-Abelian case. An important ingredient was the knowledge of the renormalization flow, to decide whether lowest-order perturbation theory could be a reliable guide to the behavior of the energy functional. Indeed, Politzer went ahead with his own calculation of the β function for Yang-Mills theory.

Our calculation proceeded slowly. I was involved in the other parts of my program and there were some tough issues to resolve. We first tried to prove on general grounds, using spectral representations and unitarity, that the theory could not be asymptotically free, generalizing the arguments of Coleman and myself to this case. This did not work, so we proceeded to calculate the β function for a Yang-Mills theory. Today this calculation is regarded as quite simple and even assigned as a homework problem in quantum field theory courses. At the time it was not so easy. This change in attitude is the analog, in theoretical physics, of the familiar phenomenon in experimental physics whereby yesterday's great discovery becomes today's background. It is always easier to do a calculation when you know what the result is and you are sure that the methods make sense. One problem we had to face was that of gauge invariance. Unlike QED, in which the charge renormalization was trivially gauge invariant (since the photon is neutral), the renormalization constants in QCD were all gauge dependent. However, the physics could not depend on the gauge. Another issue was the choice of regularization. Dimensional regularization had not really been developed yet, and we had to convince ourselves that the one-loop β function was insensitive to the regularization used. We did the calculation in an arbitrary gauge. Since we knew that the answer had to be gauge invariant, we could use gauge invariance as a check on our arithmetic. This was good since we both kept on making mistakes. In February the pace picked up, and we completed the calculation in a spurt of activity. At one point a sign error in one term convinced us that the theory was, as expected, nonasymptotically free. As I sat down to put it all together and to write up our results, I caught the error. At almost the same time Politzer finished his calculation and we compared our results. The agreement was satisfying (Gross and Wilczek, 1973b; Politzer, 1973).

Why are non-Abelian gauge theories asymptotically free? Today we can understand this in a very physical fashion, although it was certainly not so clear in 1973. It is instructive to interrupt the historical narrative and explain, in modern terms, why QCD is asymptotically free. The easiest way to understand this is by considering the magnetic screening properties of the vacuum (Nielsen, 1981). In a relativistic theory one can calculate the dielectric constant ϵ in terms of the magnetic permeability, μ , since $\epsilon\mu=1$ (in units where $c=\text{velocity of light}=1$). In classical physics all media are diamagnetic. This is because, classically, all magnets arise from electric currents and the response of a system to an applied magnetic field is to set up currents that act to decrease the field (Lenz's law). Thus $\mu<1$, a situation that corresponds to electric screening or $\epsilon>1$. However, in quantum-

mechanical systems paramagnetism is possible. This is the case in non-Abelian gauge theories where the gluons are charged particles of spin 1. They behave as permanent color magnetic dipoles that align themselves parallel to an applied external field, increasing its magnitude and producing $\mu > 1$. We can therefore regard the antiscreening of the Yang-Mills vacuum as paramagnetism. QCD is asymptotically free because the antiscreening of the gluons overcomes the screening due to the quarks. The arithmetic works as follows. The contribution to ϵ (in some units) from a particle of charge q is $-q^2/3$, arising from ordinary dielectric (or diamagnetic) screening. If the particle has spin s (and thus a permanent dipole moment γs), it contributes $(\gamma s)^2$ to μ . Thus a spin-one gluon (with $\gamma=2$, as in Yang-Mills theory) gives a contribution to μ of $\delta\mu = (-1/3 + 2^2)q^2 = 11/3q^2$; whereas a spin one-half quark contributes $\delta\mu = -[-1/3 + (2/2)^2]q^2 = -2/3q^2$ (the extra minus arises because quarks are fermions). In any case, the upshot is that as long as there are not too many quarks the antiscreening of the gluons wins out over the screening of the quarks. The formula for the β function of a non-Abelian gauge theory is given by

$$\beta(\alpha) \equiv \mu \frac{d\alpha(\mu)}{d\mu} = \frac{\alpha^2}{\pi} b_1 + \frac{\alpha^3}{\pi^2} b_2 + \dots, \quad (1)$$

where $\alpha = g^2/4\pi$.

Our result was that (Gross and Wilczek, 1973b; Politzer, 1973)

$$b_1 = - \left[\frac{11}{6} C_A - \frac{2}{3} \sum_R n_R T_R \right]. \quad (2)$$

Here C_R is the eigenvalue of the quadratic Casimir operator in the representation R of $SU(N)$ (for the adjoint representation, $C_A = N$), T_R is the trace of the square of the generators for the representation R of $SU(N)$ ($T_A = N$ and for the fundamental representation, $T_F = 1/2$), and n_R is the number of fermions in the representation R . In the case of $SU(3)$, as in QCD, $C_A = 3$, $T_F = 1/2$, and thus

$$b_1 = -11/2 + n_F/3. \quad (3)$$

Thus one can tolerate as many as 16 triplets of quarks before losing asymptotic freedom.

NON-ABELIAN GAUGE THEORIES OF THE STRONG INTERACTIONS

For me the discovery of asymptotic freedom was totally unexpected. Like an atheist who has just received a message from a burning bush, I became an immediate true believer. Field theory was not wrong—instead scaling must be explained by an asymptotically free gauge theory of the strong interactions. Our first paper contained, in addition to the report of the asymptotic freedom of Yang-Mills theory, the hypothesis that this could offer an explanation for scaling, a remark that there would be logarithmic violations of scaling, and most im-

portant of all the suggestion that the strong interactions must be based on a color gauge theory (Gross and Wilczek, 1973b).

Our abstract reads: “It is shown that a wide class of non-Abelian gauge theories have, up to calculable logarithmic corrections, free-field asymptotic behavior. It is suggested that Bjorken scaling may be obtained from strong-interaction dynamics based on non-Abelian gauge symmetry.” The first paragraph reads: “Non-Abelian gauge theories have received much attention recently as a means of constructing unified and renormalizable theories of the weak and electromagnetic interactions. In this note we report on an investigation of the ultraviolet asymptotic behavior of such theories. We have found that they possess the remarkable feature, perhaps unique among renormalizable theories, of asymptotically approaching free-field theory. Such asymptotically free theories will exhibit, for matrix elements of currents between on-mass-shell states, Bjorken scaling. We therefore suggest that one should look to a non-Abelian gauge theory of the strong interactions to provide the explanation for Bjorken scaling, which has so far eluded field-theoretic understanding.”

We had a specific theory in mind. Since the deep-inelastic experiments indicated that the charged constituents of the nucleon were quarks, the gluons had to be flavor neutral. Thus the gluons could not couple to flavor. We were very aware of the growing arguments for the color quantum number—not just the quark-model spectroscopy that was the original motivation of Han, Nambu, and Greenberg, but the counting factor (of 3) that went into the evaluation of the $\pi \rightarrow 2\gamma$ decay rate from the axial anomaly (this had been recently emphasized by Bardeen, Fritzsche, and Gell-Mann, 1973) and the factor of 3 that color provided in the total annihilation cross section. Thus the gluons could couple to color and all would be well. Hence we proposed (Gross and Wilczek, 1973b): “One particularly appealing model is based on three triplets of fermions, with Gell-Mann’s $SU(3) \times SU(3)$ as a global symmetry and an $SU(3)$ ‘color’ gauge group to provide the strong interactions. That is, the generators of the strong-interaction gauge group commute with ordinary $SU(3) \times SU(3)$ currents and mix quarks with the same isospin and hypercharge but different ‘color.’ In such a model the vector mesons are (flavor) neutral, and the structure of the operator-product expansion of electromagnetic or weak currents is essentially that of the free-quark model (up to calculable logarithmic corrections).” Thus we proposed that the strong interactions be described by the theory we now call QCD!

Callan and I had already discussed the appearance of logarithmic corrections to scaling in asymptotically free theories (Callan and Gross, 1973). We analyzed deep-inelastic scattering in an asymptotically free theory and discovered that “in such asymptotically free theories naive scaling is violated by calculable logarithmic terms.” Thus we were well aware what the form of the scaling deviations would be in such a theory. Wilczek and I immediately started to calculate the logarithmic deviations

from scaling. We were tremendously excited by the possibility of deriving exact experimental predictions from first principles that could conclusively test our asymptotically free theories of the strong interactions. We had already evaluated the asymptotic form of the flavor non-singlet structure functions, which were the easiest to calculate, at the time our *Physical Review Letter* (Gross and Wilczek, 1973b) was written, but did not have room to include the results. We immediately started to write a longer paper in which the structure of the theory would be spelled out in more detail and the dynamical issues would be addressed, especially the issue of confinement. In our letter we were rather noncommittal on this issue. We had tentatively concluded that Higgs mesons would destroy asymptotic freedom, but had only begun to explore the dynamical consequences of unbroken color symmetry. The only thing we were sure of was that “perturbation theory is not trustworthy with respect to the stability of the symmetric theory nor to its particle content” (Gross and Wilczek, 1973b). Politzer’s paper appeared just after ours. He pointed out the asymptotic freedom of Yang-Mills theory and speculated on its implications for the dynamical symmetry breaking of these theories. His abstract reads, “An explicit calculation shows perturbation theory to be arbitrarily good for the deep Euclidean Green’s functions of any Yang-Mills theory and of many Yang-Mills theories with fermions. Under the hypothesis that spontaneous symmetry breakdown is of dynamical origin, these symmetric Green’s functions are the asymptotic forms of the physically significant spontaneously broken solution, whose coupling could be strong.” No mention is made of either Bjorken scaling or of the strong interactions (Politzer, 1973).

In our second paper, written a few months later, we outlined in much greater detail the structure of asymptotically free gauge theories of the strong interactions and the predictions for the scaling violations in deep-inelastic scattering (Gross and Wilczek, 1973a). The paper was delayed for about two months because we had problems with the singlet structure functions—due to the operator mixing of physical operators with ghost operators. This problem was similar to the issue of gauge invariance that had plagued us before. Here the problem was more severe. Physical operators, whose matrix elements were measurable in deep-inelastic scattering experiments, mixed under renormalization with ghost operators that could have no physical meaning. Finally we deferred the analysis of the singlet structure functions to a third paper, in which we resolved this issue (Gross and Wilczek, 1974). We showed that, even though this mixing was real and unavoidable, the ghost operators decoupled from physical measurements. In the second paper we discussed in detail the choice between symmetry breaking and unbroken symmetry and noted that “Another possibility is that the gauge symmetry is exact. At first sight this would appear to be ridiculous since it would imply the existence of massless, strongly coupled vector mesons. However, in asymptotically free theories these naive expectations might be wrong. There may be

little connection between the ‘free’ Lagrangian and the spectrum of states.... The infrared behavior of Green’s functions in this case is determined by the strong-coupling limit of the theory. It may be very well that this infrared behavior is such so as to suppress all but color singlet states, and that the colored gauge fields as well as the quarks could be ‘seen’ in the large-Euclidean-momentum region but never produced as real asymptotic states” (Gross and Wilczek, 1973a).

Steve Weinberg reacted immediately to asymptotic freedom. He wrote a paper in which he pointed out that in an asymptotically free gauge theory of the strong interactions the order- α interactions produced by electroweak interactions can be calculated ignoring the strong force, and he found that these effects do not violate conservation of parity and strangeness, in agreement with observation, as long as there were no colored scalars (Weinberg, 1973). This led him to suggest that a theory with unbroken color symmetry could explain why we do not see quarks and gluons. There is a slight difference between our respective conjectures. Weinberg argued that perhaps the infrared divergences, caused by the masslessness of the gluons in an unbroken color gauge theory, would make the rate of production of non-singlet states vanish. Today we believe in the existence of nonconfining, Coulomb phases, with unbroken color symmetry, for some supersymmetric non-Abelian gauge theories. We argued that perhaps the growth of the effective coupling at large distances, the infrared behavior of the coupling caused by the flip side of asymptotic freedom (later dubbed infrared slavery by Georgi and Glashow), would confine the quarks and gluons in color singlet states.

In October 1973 Fritsch, Gell-Mann, and H. Leutwyler submitted a paper in which they discussed the “advantages of the color octet gluon picture.” Here they discussed the advantages of “abstracting properties of hadrons and their currents from a Yang-Mills gauge model based on colored quarks and color octet gluons” (Fritsch *et al.*, 1973). They discussed various models and pointed out the advantages of each. The first point was already discussed at the National Accelerator Laboratory high-energy physics conference in August 1972. There Gell-Mann and Fritsch had discussed their program of “abstracting results from the quark-gluon model” (Fritsch and Gell-Mann, 1973). They discussed various models and asked “We shall treat the vector gluon, for convenience, as a color singlet”? In October 1973 Fritsch, Gell-Mann, and Leutwyler also noted that in the nonrelativistic quark model with a Coulomb potential mediated by vector gluons the potential is attractive in color singlet channels, which might explain why these are light, a point that had been made previously by H. Lipkin (1973). They also noted the asymptotic freedom of such theories, but did not regard this as an argument for scaling since “we conjecture that there might be a modification at high energies that produces true scaling.” Finally they noted that the axial U(1) anomaly in a non-Abelian gauge theory might explain the notorious U(1) problem, although they could not explain

how, since the anomaly itself could be written as a total divergence. [It required the discovery of instantons to find the explanation of the U(1) problem.]

THE EMERGENCE AND ACCEPTANCE OF QCD

Although it was clear to me that the strong interactions must be described by non-Abelian gauge theories, there were many problems. The experimental situation was far from clear, and the issue of confinement remained open. However, within a small community of physicists the acceptance of the theory was very rapid. New ideas in physics sometimes take years to percolate into the collective consciousness. However, in rare cases such as this there is a change of perception analogous to a phase transition. Before asymptotic freedom it seemed that we were still far from a dynamical theory of hadrons; afterwards it seemed clear that QCD was such a theory. [The name quantum chromodynamics, or QCD, first appeared in a review by W. Marciano and H. Pagels (1978), where it was attributed to Gell-Mann. It was such an appropriate name that no one could complain.] Asymptotic freedom explained scaling at short distances and offered a mechanism for confinement at large distance. Suddenly it was clear that a non-Abelian gauge theory was consistent with everything we knew about the strong interactions. It could encompass all the successful strong-interaction phenomenology of the past decade. Since the gluons were flavor neutral, the global flavor symmetries of the strong interactions, $SU(3) \times SU(3)$, were immediate consequences of the theory, as long as the masses of the quarks (the mass parameters of the quarks in the Lagrangian, not the physical masses that are effectively infinite due to confinement) are small enough. Even more alluring was the fact that one could calculate. Since perturbation theory was trustworthy at short distances many problems could be tackled. Some theorists were immediately convinced, among them Altarelli, Appelquist, Callan, Coleman, Gaillard, Gatto, Georgi, Glashow, Kogut, Lee, Maiani, Migdal, Polyakov, Politzer, Susskind, Weinberg, and Zee. At large distances, however, perturbation theory was useless. In fact, even today after 31 years of study we still lack reliable, analytic tools for treating this region of QCD. This remains one of the most important areas of theoretical particle physics. However, at the time the most important thing was to convince oneself that the idea of confinement was not inconsistent. One of the first steps in that direction was provided by lattice gauge theory. I first heard of Wilson's lattice gauge theory (Wilson, 1974) when I gave a lecture at Cornell in the late spring of 1973. Wilson had started to think of this approach soon after asymptotic freedom was discovered. The lattice formulation of gauge theory (independently proposed by Polyakov) had the enormous advantage, as Wilson pointed out in the fall of 1973, that the strong-coupling limit was particularly simple and exhibited confinement. Thus one had at least a crude approximation in which confinement was exact. It is a very crude approximation, since to arrive at the continuum theory

from the lattice theory one must take the weak-coupling limit. However, one could imagine that the property of confinement was not lost as one went continuously from strong to weak lattice coupling, i.e., there was no phase transition. Moreover, one could, as advocated by Wilson, study this possibility numerically using Monte Carlo methods to construct the lattice partition function. However, the first quantitative results of this program did not emerge until 1981. By now the program of calculating the hadronic mass spectrum has come close to its goal, achieving reliable results that fit the low-lying spectrum to a few percent!

Personally I derived much solace in the coming year from two examples of soluble two-dimensional field theories. One was the $(\bar{\Psi}\Psi)^2$ theory that Neveu and I analyzed and solved for large N (Gross and Neveu, 1974). This provided a soluble example of an asymptotically free theory that underwent dimensional transmutation, solving its infrared problems by generating a dynamical fermion mass through spontaneous symmetry breaking. This provided a model of an asymptotically free theory, with no built-in mass parameters. We could solve this model and check that it was consistent and physical. The other soluble model was two-dimensional QCD, analyzed by 't Hooft in the large N limit ('t Hooft, 1974). Two-dimensional gauge theories trivially confine color. This was realized quite early and discussed for Abelian gauge theory, the Schwinger model, by Casher, Kogut, and Susskind (1973, 1974) as a model for confinement in the fall of 1973. However QCD₂ is a much better example. It has a spectrum of confined quarks which in many ways resembles the four-dimensional world. These examples gave many of us total confidence in the consistency of the concept of confinement. It clearly was possible to have a theory whose basic fields do not correspond to asymptotic states, to particles that one can observe directly in the laboratory. Applications of the theory also began to appear. Two calculations of the β function to two-loop order were performed, with the result that in the notation of Eq. (2),

$$b_2 = - \left[\frac{17}{12} C_A^2 - \frac{1}{2} C_F T_F n - \frac{5}{6} C_A T_F n \right]$$

(Caswell, 1974; Jones, 1974). Appelquist and Georgi (1973) and Zee (1973b) calculated the corrections to the scaling of the e^+e^- annihilation cross section; Gaillard and Lee (1974), and independently Altarelli and Maiani (1974), calculated the enhancement of the $\Delta I=1/2$ non-leptonic decay matrix elements. The analysis of scaling violations for deep-inelastic scattering continued (Gross, 1974), and the application of asymptotic freedom, what is now called perturbative QCD, was extended to many new processes.

The experimental situation developed slowly and initially looked rather bad. I remember in the spring of 1974 attending a meeting in Trieste. There I met Burt Richter, who was gloating over the fact that $R = \sigma_{e^+e^- \rightarrow \text{hadrons}} / \sigma_{e^+e^- \rightarrow \mu^+\mu^-}$ was increasing with energy, instead of approaching the expected constant value. This

was the most firm of all the scaling predictions. R must approach a constant in any scaling theory. In most theories one cannot predict the value of the constant. However, in an asymptotically free theory the constant is predicted to equal the sum of the squares of the charges of the constituents. Therefore, if there were only the three observed quarks, one would expect that $R \rightarrow 3[(1/3)^2 + (1/3)^2 + (2/3)^2] = 2$. However, Richter reported that R was increasing, passing through 2, with no sign of flattening out. Now many of us knew that charmed particles had to exist. Not only were they required, indeed invented, for the Glashow-Illiopoulos-Maiani mechanism to work, but as C. Bouchiat, J. Illiopoulos, and P. Meyer (1972), and independently R. Jackiw and I (Gross and Jackiw, 1972) showed, if the charmed quark were absent the electroweak theory would be anomalous and non-renormalizable. Gaillard, Lee, and Rosner had written an important and insightful paper on the phenomenology of charm (Gaillard *et al.*, 1975). Thus many of us thought that, since R was increasing, probably charm was being produced. In 1974 the charmed mesons, much narrower than anyone imagined [except for Appelquist and Politzer (1975)], were discovered, looking very much like positronium and easily interpreted as Coulomb bound states of quarks. This clinched the matter for many of the remaining skeptics. The rest were probably convinced once experiments at higher energy began to see quark and gluon jets.

The precision tests of the theory, the logarithmic deviations from scaling, took quite a while to observe. I remember very well a remark made to me by a senior colleague, in April of 1973 when I was very excited, right after the discovery of asymptotic freedom. He remarked that it was unfortunate that our new predictions regarding deep-inelastic scattering were logarithmic effects, since it was unlikely that we would see them verified, even if true, in our lifetime. This was an exaggeration, but the tests did take a long time to appear. Confirmation only started to trickle in between 1975 and 1978 at a slow pace. By now the predictions are indeed verified, in many cases to better than 1%. Nowadays, when you listen to experimentalists talk about their results they point to their Lego plots and say, "Here we see a quark, here a gluon." Believing is seeing, seeing believes. We now believe in the physical reality of quarks and gluons; we now believe in the asymptotic simplicity of their interactions at high energies, so we can see quarks and gluons. The way in which we see quarks and gluons, indirectly through the effects they have on our measuring instruments, is not much different from the way we see electrons.

IMPLICATIONS OF ASYMPTOTIC FREEDOM

The most important implication of asymptotic freedom is QCD itself with pointlike behavior of quarks at short distance and the strong confining force at large distance. But in addition, asymptotic freedom greatly increased our confidence in the consistency of quantum field theory, produced the first example of a theory with

no adjustable parameters, enabled us to probe the very early history of the universe, and allowed us to extrapolate the Standard Model to high energy.

A. Consistency of quantum field theory

Traditionally, fundamental theories of nature have had a tendency to break down at short distances. This often signals the appearance of new physics that is discovered once one has experimental instruments of high enough resolution (energy) to explore the higher-energy regime. Before asymptotic freedom it was expected that any quantum field theory would fail at sufficiently high energy, where the flaws of the renormalization procedure would appear. To deal with this, one would have to invoke some kind of fundamental length. In an asymptotically free theory this is not necessarily the case; the decrease of the effective coupling for large energy means that no new physics need arise at short distances. There are no infinities at all, the bare coupling is finite, and in fact it vanishes. The only divergences that arise are an illusion that appears when one tries to compare, in perturbation theory, the finite effective coupling at finite distances with the vanishing effective coupling at infinitely short distances.

Thus the discovery of asymptotic freedom greatly reassured us of the consistency of four-dimensional quantum field theory. We can trust renormalization theory in asymptotically free theories, even though perturbation theory is only an asymptotic expansion, since it gets simpler in the regime of short distances. We are very close to having a rigorous mathematical proof of the existence of asymptotically free gauge theories in four dimensions—at least when placed into a finite box to tame the infrared dynamics that produces confinement.

B. No adjustable parameters

At first glance QCD has only one parameter, the dimensionless number that specifies the strength of the force (if we neglect the quark masses, an excellent approximation for ordinary hadrons since the light quarks are so light). But through the dependence of the charge on distance or energy, the theory produces a dynamical mass scale. One defines the mass scale of QCD to be the energy at which the charge equals some value, say 1. Then, via this phenomenon of dimensional transmutation, all masses, indeed all observables, are calculable in terms of the dynamically generated mass scale. It is sometimes claimed that the origin of mass is the Higgs mechanism that is responsible for the breaking of the electroweak symmetry that unbroken would forbid quark masses. This is incorrect. Most, 99%, of the proton mass is due to the kinetic and potential energy of the massless gluons and the essentially massless quarks, confined within the proton.

Thus QCD provides the first example of a complete theory, with no adjustable parameters and with no indication within the theory of a distance scale at which it must break down. Indeed, were it not for the elec-

troweak interactions and gravity, we might be satisfied with QCD as it stands. It is the best example we possess of a perfect, complete theory.

C. The early history of the universe

The universe has been expanding since the big bang; thus early on it was hot and dense. To trace the history of the universe we must understand the dynamics that operated when the universe was hot and particles were very energetic. Before the Standard Model we could not go back further than 200 000 years after the big bang. Today, especially since QCD simplifies at high energy, we can extrapolate to very early times when nucleons melted and quarks and gluons were liberated to form a quark-gluon plasma.

D. Unification

One of the most important implications of asymptotic freedom is the insight it gave into the unification of all the forces of nature. Almost immediately after the discovery of asymptotic freedom and the proposal of the non-Abelian gauge theories of the strong interactions, the first attempts were made to unify all the interactions. This was natural, given that one was using very similar theories to describe all the known interactions. Furthermore, the apparently insurmountable barrier to unification—namely, the large difference in the strengths of the strong interactions and the electroweak interactions—was seen to be a low-energy phenomenon. Since the strong interactions decrease in strength with increasing energy, these forces could have a common origin at very high energy. Georgi, Quinn, and Weinberg (1974) showed that the couplings run in such a way as to merge somewhere around 10^{14} – 10^{16} GeV. This is our most direct clue as to where the next threshold of fundamental physics lies, and it hints that at this immense energy all the forces of nature, including gravity, are unified.

ACKNOWLEDGMENTS

As I end I would like to thank not only the Nobel Foundation, but nature itself, who has given us the opportunity to explore her secrets and the fortune to have revealed one of her most mysterious and beautiful aspects—the strong force.

REFERENCES

Adler, S., 1965, “Sum rules for axial vector coupling constant renormalization in beta decay,” *Phys. Rev.* **140**, B736–B747.
 Altarelli, G., and L. Maiani, 1974, “Octet enhancement of non-leptonic weak interactions in asymptotically free gauge theories,” *Phys. Lett.* **52B**, 351–354.
 Appelquist, T., and H. Georgi, 1973, “ e^+e^- annihilation in gauge theories of strong interactions,” *Phys. Rev. D* **8**, 4000–4002.
 Appelquist, T., and H. D. Politzer, 1975, “Heavy quarks and

e^+e^- annihilation,” *Phys. Rev. Lett.* **34**, 43–45.
 Bardeen, W. A., H. Fritzsch, and M. Gell-Mann, 1973, “Light-cone current algebra, π^0 decay, and e^+e^- annihilation,” in *Scale and Conformal Symmetry in Hadron Physics*, edited by R. Gatto (Wiley, New York).
 Bjorken, J. D., 1966, “Applications of the chiral $U(6)\times U(6)$ algebra of current densities,” *Phys. Rev.* **148**, 1467–1478.
 Bjorken, J. D., 1968, “Current algebra at small distances,” *Proceedings of the International School of Physics*, “Enrico Fermi,” Course XLI, Varenna School Lectures, Varenna Italy, 1967, edited by J. Steinberger (Academic, New York).
 Bloom, E. D., *et al.*, 1969, “High-energy inelastic $e-p$ scattering at 6° and 10° ,” *Phys. Rev. Lett.* **23**, 930–934.
 Bogoliubov, N. N., and D. V. Shirkov, 1959, *Introduction to the Theory of Quantized Fields* (Interscience, New York).
 Bouchiat, C., J. Iliopoulos, and Ph. Meyer, 1972, “An anomaly free version of Weinberg’s model,” *Phys. Lett.* **38B**, 519–523.
 Callan, C. G., 1970, “Broken scale invariance in scalar field theory,” *Phys. Rev. D* **2**, 1541–1547.
 Callan, C. G., 1972, “Broken scale invariance and asymptotic behavior,” *Phys. Rev. D* **5**, 3202–3210.
 Callan, C. G., and D. J. Gross, 1968a, “Crucial test of a theory of currents,” *Phys. Rev. Lett.* **21**, 311–313.
 Callan, C. G., and D. J. Gross, 1968b, “High-energy electroproduction and the constitution of the electric current,” *Phys. Rev. Lett.* **22**, 156–159.
 Callan, C. G., and D. J. Gross, 1973, “Bjorken scaling in quantum field theory,” *Phys. Rev. D* **8**, 4383–4394.
 Casher, A., J. Kogut, and L. Susskind, 1973, “Vacuum polarization and the quark-parton puzzle,” *Phys. Rev. Lett.* **31**, 792–795.
 Casher, A., J. Kogut, and L. Susskind, 1974, “Vacuum polarization and the absence of free quarks,” *Phys. Rev. D* **10**, 732–745.
 Caswell, W., 1974, “Asymptotic behavior of non-Abelian gauge theories to two-loop order,” *Phys. Rev. Lett.* **33**, 244–246.
 Chew, G., 1963, *S-Matrix Theory of Strong Interactions* (Benjamin, New York).
 Christ, N., B. Hasslacher, and A. Mueller, 1972, “Light-cone behavior of perturbation theory,” *Phys. Rev. D* **6**, 3543–3562.
 Coleman, S., and D. J. Gross, 1973, “Price of asymptotic freedom,” *Phys. Rev. Lett.* **31**, 851–854.
 Dashen, R., and M. Gell-Mann, 1966, “Representations of local current algebra at infinite momentum,” *Phys. Rev. Lett.* **17**, 340–343.
 Drell, S. D., and T. M. Yan, 1971, “Partons and their applications at high energies,” *Ann. Phys. (N.Y.)* **66**, 578–623.
 Feynman, R. P., 1962, in *The Quantum Theory of Fields*, Proceedings of the Twelfth Conference on Physics at the University of Brussels, October, 1961 (Interscience, New York).
 Feynman, R. P., 1969, “High-energy collisions of hadrons,” *Phys. Rev. Lett.* **23**, 1415–1417.
 Frishman, Y., 1971, “Operator products at almost light like distances,” *Ann. Phys. (N.Y.)* **66**, 373–389.
 Fritzsch, H., and M. Gell-Mann, 1973, “Current algebra: Quarks and what else?” in *Proceedings of the XVI International Conference on High Energy Physics*, The University of Chicago and National Accelerator Laboratory, 1972, edited by J. D. Jackson and A. Roberts (National Accelerator Laboratory, Batavia, IL), Vol. 2, p. 164.
 Fritzsch, H., M. Gell-Mann, and H. Leutwyler, 1973, “Advan-

- tages of the color octet gluon picture," *Phys. Lett.* **47B**, 365–368.
- Gaillard, M. K., and B. W. Lee, 1974, " $\Delta I = \frac{1}{2}$ rule for non-leptonic decays in asymptotically free field theories," *Phys. Rev. Lett.* **33**, 108–111.
- Gaillard, M. K., B. W. Lee, and J. L. Rosner, 1975, "Search for charm," *Rev. Mod. Phys.* **47**, 277–310.
- Gell-Mann, M., 1964a, "The symmetry group of vector and axial vector currents," *Physics (Long Island City, N.Y.)* **1**, 63.
- Gell-Mann, M., 1964b, "A schematic model of baryons and mesons," *Phys. Lett.* **8**, 214.
- Gell-Mann, M., 1973, "Current algebra: Quarks and what else?" in *Proceedings of the XVI International Conference on High Energy Physics*, The University of Chicago and National Accelerator Laboratory, 1972, edited by J. D. Jackson and A. Roberts (National Accelerator Laboratory, Batavia, IL), Vol. 4, p. 135.
- Gell-Mann, M., and F. Low, 1954, "Quantum electrodynamics at small distances," *Phys. Rev.* **95**, 1300–1312.
- Gell-Mann, M., and Y. Neeman, 1964, *The Eightfold Way* (Benjamin, New York).
- Georgi, H., H. R. Quinn, and S. Weinberg, 1974, "Hierarchy of interactions in unified gauge theories," *Phys. Rev. Lett.* **33**, 451–454.
- Greenberg, O. W., 1964, "Spin and unitary-spin independence in a paraquark model of baryons and mesons," *Phys. Rev. Lett.* **13**, 598–602.
- Gross, D. J., 1971, "Light cone structure of current commutators in the gluon-quark model," *Phys. Rev. D* **4**, 1059–1072.
- Gross, D. J., 1974, "How to test scaling in asymptotically free theories," *Phys. Rev. Lett.* **32**, 1071–1073.
- Gross, D. J., and R. Jackiw, 1972, "The effect of anomalies on quasi-renormalizable theories," *Phys. Rev. D* **6**, 477–493.
- Gross, D. J., and C. Llewellyn-Smith, 1969, "High energy neutrino-nucleon scattering, current algebra and partons," *Nucl. Phys. B* **14**, 337–347.
- Gross, D. J., and A. Neveu, 1974, "Dynamical symmetry breaking in an asymptotically free theory," *Phys. Rev. D* **10**, 3235–3253.
- Gross, D. J., and J. Wess, 1970, "Scale invariance, conformal invariance and the high energy behavior of scattering amplitudes," *Phys. Rev. D* **2**, 753–764.
- Gross, D. J., and F. Wilczek, 1973a, "Asymptotically free gauge theories. I," *Phys. Rev. D* **8**, 3633–3652.
- Gross, D. J., and F. Wilczek, 1973b, "Ultraviolet behavior of non-Abelian gauge theories," *Phys. Rev. Lett.* **30**, 1343–1346.
- Gross, D. J., and F. Wilczek, 1974, "Asymptotically free gauge theories. II," *Phys. Rev. D* **9**, 980–993.
- Han, M. Y., and Y. Nambu, 1965, "Three-triplet model with double SU(3) symmetry," *Phys. Rev.* **139**, B1006–B1010.
- Jackiw, R., R. Van Royen, and G. West, 1970, "Measuring light cone singularities," *Phys. Rev. D* **2**, 2473–2485.
- Jones, D., 1974, "Two-loop diagrams in Yang-Mills theory," *Nucl. Phys. B* **75**, 531–538.
- Landau, L. D., 1955, "On the quantum theory of fields," in *Niels Bohr and the Development of Physics*, edited by W. Pauli (Pergamon, London), p. 52.
- Landau, L. D., 1960, "Fundamental problems," in *Theoretical Physics in the Twentieth Century: A Memorial Volume to Wolfgang Pauli*, edited by M. Fierz and V. F. Weisskopf (Interscience, New York), p. 245.
- Landau, L. D., and I. Pomeranchuk, 1955, "Radiation of γ -rays upon collision of fast π -mesons with nucleons," *Dokl. Akad. Nauk SSSR* **102**, 489.
- Leutwyler, H., and J. Stern, 1970, "Singularities of current commutators on the light cone," *Nucl. Phys. B* **20**, 77–101.
- Lipkin, H., 1973, "Triality, exotics and the dynamical basis of the quark model," *Phys. Lett.* **45B**, 267–271.
- Low, F., 1967, *Proceedings of the International Conference on High Energy Physics 1966* (University of California, Berkeley), p. 249.
- Marciano, W., and H. Pagels, 1978, "Quantum chromodynamics," *Phys. Rep.* **36**, 137–276.
- Nambu, Y., 1965, "Dynamical symmetries & fundamental fields," in *Proceedings of the Second Coral Gables Conference on Symmetry Principles at High Energy* (Freeman, San Francisco), p. 133.
- Nambu, Y., 1966, "Systematics of hadrons in subnuclear physics," in *Preludes in Theoretical Physics: In Honor of V. F. Weisskopf*, edited by A. de Shalit, H. Feshbach, and L. van Hove (North-Holland, Amsterdam).
- Nielsen, N. K., 1981, "Asymptotic freedom as a spin effect," *Am. J. Phys.* **49**, 1171–1178.
- Parisi, G., 1973, "Deep inelastic scattering in a field theory with computable large-momenta behavior," *Lett. Nuovo Cimento Soc. Ital. Fis.* **7**, 84.
- Politzer, H. D., 1973, "Reliable perturbative results for strong interactions?" *Phys. Rev. Lett.* **30**, 1346–1349.
- Stueckelberg, E., and A. Petermann, 1953, "La normalisation des constantes dans la theorie des quanta," *Helv. Phys. Acta* **26**, 499.
- Symansik, K., 1970, "Small distance behavior in field theory and power counting," *Commun. Math. Phys.* **18**, 227.
- Symansik, K., 1971, "Small-distance-behavior analysis and Wilson expansions," *Commun. Math. Phys.* **23**, 49.
- 't Hooft, G., 1967, "Renormalizable Lagrangians for massive Yang-Mills fields," *Nucl. Phys. B* **35**, 167–188.
- 't Hooft, G., 1974, "A planar diagram theory for strong interactions," *Nucl. Phys. B* **72**, 461–473.
- Terent'ev, M. V., and V. S. Vanyashin, 1965, "The vacuum polarization of a charged vector field," *Sov. Phys. JETP* **21**, 380.
- Veneziano, G., 1968, "Construction of a crossing-symmetric Regge-behaved amplitude for linearly rising trajectories," *Nuovo Cimento A* **57**, 190.
- Weinberg, S., 1973, "Non-Abelian gauge theories of the strong interactions," *Phys. Rev. Lett.* **31**, 494–497.
- Weisberger, W. I., 1966, "Unsubtracted dispersion relations and the renormalization of the weak axial-vector coupling constants," *Phys. Rev.* **143**, 1302–1309.
- Wilson, K., 1971, "Renormalization group and strong interactions," *Phys. Rev. D* **3**, 1818–1846.
- Wilson, K., 1974, "Confinement of quarks," *Phys. Rev. D* **10**, 2445–2459.
- Wilson, K., and J. Kogut, 1974, "The renormalization group and the ϵ expansion," *Phys. Rep., Phys. Lett.* **12**, 75–199.
- Zee, A., 1973a, "Study of the renormalization group for small coupling constants," *Phys. Rev. D* **7**, 3630–3636.
- Zee, A., 1973b, "Electron-positron annihilation in stagnant field theories," *Phys. Rev. D* **8**, 4038–4041.
- Zweig, G., 1964, CERN Report No. TH401, 4R12 (unpublished).