Introduction to QCD

http://www.nobelprize.org/mediaplayer/index.php?id=569 Wilczek 35 minutes

http://www.nobelprize.org/mediaplayer/index.php?id=566 Gross 44 minutes

Wilczek at Berkeley
http://www.youtube.com/watch?v=914jzZ4LXcU&feature=BFa&list=SP76C921BD8CBD0B8E

Gross at Berkeley
http://www.youtube.com/watch?v=Mo05DBiCrLc&feature=BFa&list=SP76C921BD8CBD0B8E

Frank Wilczek on QCD: An Unfinished Symphony
http://www.youtube.com/watch?v=pkmTmCVzZJs

Feynman diagrams and gauge fields
http://www.youtube.com/watch?v=4TX7CcAPF44 6 1/2 minutes

Wilczek Interview
http://www.youtube.com/watch?v=CQ-HjjV7KPI

THE DILEMMA OF ATTRIBUTION

Nobel Lecture, December 8, 2004

by

H. David Politzer

California Institute of Technology (Caltech), Pasadena, USA.

I suspected that there were some members of the live audience who were somewhat apprehensive about sitting through the morning's physics lectures. After all, there were three guys there to talk about one minus sign. If it were just two people and a plus sign, +, one could talk about the | and the other about the –. However, to my mind, this year's awards represent or symbolize not just a minus sign but a large body of significant advances in our understanding of fundamental physics and are the work of not just three people but a great many scientists, stretching out over many years and many countries. This is really a prize for that whole community.

Sidney Coleman, my beloved teacher from graduate school, referred to this community as *i fratelli fisici,* by which he meant the brotherhood of physicists. Most of us spoke at least a bit of broken Italian, a legacy of the grand and highly influential summer schools organized by Nino Zichichi in Erice, Sicily. Indeed, one of my fondest reflections on my particle physics career is having been able to arrive at a train station, virtually anywhere in the world, and be greeted by a total stranger who immediately treated me like an old friend.

I'd love to tell you all their stories, but I certainly don't know them all, nor do I have time (or space) even for those that I do. So I've chosen a few of the people and a few of the stories with which to make a particular point. You can judge for yourself at the end how well I've succeeded. And I'll deal mostly with theorists because I know them best – although I must say that I do regard theoretical physics as a fundamentally parasitic profession, living off the labors of the real physicists.

I'd like to address one particular aspect of the impact of these prizes. To a considerable extent they have come to represent milestones in the progress of science. And it is a testament to the care and wisdom exercised in the selection process just how important the prizes have become. To the public, they spark continued interest in science's most important advances. But even within the world of the scientific experts, the prizes likewise serve as markers of this progress. The use of history in science education may be a contributing factor to why this is so and how it works. As teachers of the next generation of scientists, we always seek to compress and simplify all the developments that have come before. We want to bring our students as quickly as possible to the frontier of current understanding. From this perspective, the actual history, which involves many variants and many missteps, is only a hindrance. And the neat, linear progress, as outlined by the sequence of gleaming gems recognized by Nobel Prizes, is a useful fiction. But a fiction it is. The truth is often far more complicated. Of course, there are the oft-told priority disputes, bickering over who is responsible for some particular idea. But those questions are not only often unresolvable, they are often rather meaningless. Genuinely independent discovery is not only possible, it occurs all the time. Sometimes a yet harder problem in the prize selection process is to identify what is the essential or most important idea in some particular, broader context. So it's not just a question of who did it, i.e., who is responsible for the work, but what "it" is. I.e., what is the significant "it" that should stand as a symbol for a particularly important advance.

I've no interest in recounting my whole life's story or even my physics career. Rather, I want to focus on the context of the particular work cited in this year's awards. So I begin this saga with a trip I took with Erick Weinberg, a fellow graduate student, friend, and something of a mentor (he was a year ahead of me) from Cambridge, Massachusetts to Hoboken, New Jersey (I think it was 1970) to a conference to hear our teacher, Sidney Coleman, speak. He was delivering a paper titled "Why Dilatation Generators Don't Generate Dilatations." We had read a written version but hoped that his talk would help us understand it better. It was a several hour drive. Somewhere along the way, I asked Erick to explain to me a bit about what were called Yang-Mills or non-Abelian gauge theories. I had heard the name but was otherwise ignorant. They'd been invented in 1954 and were the last and least understood entry in a short list of what came to be considered the only possible descriptions of fundamental particle interactions. Erick explained the defining basics but told me that nothing was known about their consequences and that many of the most famous senior particle theorists had gotten seriously confused about them. (The list of such notables included Dick Feynman, Shelly Glashow, Abdus Salam, and Steve Weinberg.) And now it seemed that no senior physicist wanted to discuss them; their ignorance and confusion were too embarrassing.

(While delivering my talk live in Stockholm, it occurred to me I should have had a little light or a bell that went off when I mentioned a Nobel Laureate – because part of my point is to try to understand who is and who isn't. The relevant names are already familiar to the physicist segment of the audience. But for the sake of the general audience, I just raised my finger discretely. Here I'll use a superscript $\mathcal{N}$. So far, there's Yang$^{\mathcal{N}}$, Feynman$^{\mathcal{N}}$, Glashow$^{\mathcal{N}}$, Salam$^{\mathcal{N}}$, Steve Weinberg$^{\mathcal{N}}$, but not Coleman or Erick Weinberg.)

It turns out there was one brave soul, Tini Veltman$^{\mathcal{N}}$, who never gave up on Yang-Mills theory, and, with his best-ever grad student, Gerard 't Hooft$^{\mathcal{N}}$, cracked the case in 1971. I think it worth noting that I, personally, know of no one who claimed to understand the details of 't Hooft 's paper. Rather we all learned it from Ben Lee, who combined insights from his own work (that renormalization constants are independent of the choice of ground state in such theories), from hitherto unnoticed work from Russia (Fadde'ev and Popov on quantization and Feynman rules), and from the simple encouragement from 't Hooft 's paper that it was possible. (It is amazing how much easier it can be to solve a problem once you are assured that a solution exists!)

The bit of physics I remember best from the Hoboken conference was from a talk by T.D. Lee[N]. He spoke with confidence that the weak interactions were mediated by a heavy bosonic particle that carried the force, and he gave its mass. (Several years later he was proven right.) The clearest version of that theory had been written down by Steve Weinberg[N] in 1967. But no one in that period ever referred to Weinberg's paper. For example, I don't think that Weinberg's paper had any influence on T.D. Lee's thinking. In fact, when what is now known as the Weinberg-Salam model was recognized by the Nobel Foundation, Sidney Coleman published in *Science* magazine in 1979 a citation search he did documenting that essentially no one paid any attention to Weinberg's Nobel Prize winning paper until the work of 't Hooft[N] (as explicated by Ben Lee). In 1971 interest in Weinberg's paper exploded. I had a parallel personal experience: I took a one-year course on weak interactions from Shelly Glashow[N] in 1970, and he never even mentioned the Weinberg-Salam model or his own contributions to the theory (for which he shared that Nobel Prize; by the way, his contribution to that theory was largely his PhD thesis work, done under the direction of Julian Schwinger[N], who had already published papers on the non-Abelian gauge bosons as carriers of the weak force in the mid- 1960's.) I note again that I also don't personally know anyone who ever read Salam's work on the subject either, except for John Ward, and he was actually the co-author on the relevant papers. – He is not a Nobel Laureate.

A further aside on the work of 't Hooft[N] and Veltman[N], whose contributions were enormously profound and influential, albeit really rather difficult to characterize for a lay audience. One of their many contributions (called, in the business, dimensional regularization) is a tool of essential significance, both for settling issues of principle and for doing explicit calculations. Dimensional regularization also was invented independently for the same purposes and appeared in an earlier paper, now mostly forgotten, by Bollini and Giambiaggi.

Coleman's talk in Hoboken was about his then, early understanding of what came to be known as the renormalization group. His thinking was very much influenced by the independent work of Kurt Symanzik and Curt Callan. However, the undisputed champion of the renormalization group was Ken Wilson[N], (one of my all-time, absolute heroes) for which he received Nobel recognition. That a prize was given to Wilson[N] and Wilson[N] alone in 1982 perhaps reflects the depth of his understanding, the precision of his detailed, physical predictions, and his evangelical zeal. We should remember, however, that the renormalization group work that led to experimentally confirmed predictions, which were in the field of phase transitions and are the substance of the citation for that prize, was all done in collaboration with Michael Fisher; we should remember that the basic, formal work was done independently and published earlier by Wegner and Houghton; and we should remember that the essential physical ideas were articulated independently and earlier by Leo Kadanoff. Furthermore, the renormalization group was actually invented in 1954 by Murray Gell-Mann[N] and Francis Low. But even that formulation of the renormalization group appeared in an earlier, independent paper of Stueckelberg and Petermann.

In the early days following the triumph of the Weinberg-Salam model, at one point Glashow[N] asked Coleman a practical question that came up in his own work. (The specific technical question was "What happens if the whole theory has less symmetry than the classical scalar (spin zero) sector?") And Coleman answered the question, but he also recognized that the answer was worthy of a deeper, clearer understanding. So, he embarked on its study, in the simplest possible contexts, with my buddy Erick Weinberg. I tagged along in this effort and occasionally made some contribution.

(Here's an anecdote of my first meeting with Niccolo Cabbibo, a charming man, responsible for a monumental contribution to our understanding of the weak interactions and their relation to the strong interactions, which is now largely overlooked because of the telescoping of history into a compact introduction to the present. We were both visiting the University of Chicago, staying at the Windemere Hotel. We chatted over dinner and after as rats scurried between our feet. He is the only person who ever mentioned to me noticing my name in the acknowledgments of Coleman and Weinberg's classic paper.)

During this work with Coleman and Weinberg, one day I wondered and then asked Coleman, "What happens if there are no scalar fields (spin zero particles) in the first place?" It was an innocent but inordinately profound question which occupied us both quite intensively for the next several months. I learned an enormous amount just working on it. And I benefited from far closer and more extensive interactions with Coleman than he awarded to most of his students – because he was actively working on the problem with me. However, I never made what Coleman considered substantial "progress" as measured by his standards. On the other hand, I did many things that, in retrospect, would have been publishable on their own. For example, I was very proud of a trick I invented (only to be told later that it was first done by Heisenberg[N]) for solving (at least in the simplest approximation [what's called 1/N]) what came later to be known as the Gross-Neveu model.

Coleman took a leave of absence from Harvard, taking his sabbatical at Princeton. At that point, I decided I needed a research program on which I could proceed on my own – something that might not meet Coleman's high standards but on which I might have some chance of success. I decided to look into whether the renormalization group had anything to say about the low energy (or ground state) behavior of Yang-Mills theory. An analogous analysis for electrodynamics appeared in the classic textbook of Bogoliubov and Shirkov, though Coleman characterized the relevant chapter there as "mysterious". This was a possible approach to the question I articulated regarding no scalar fields, but I thought I might be able to follow the steps of Bogoliubov and Shirkov explicitly as a guide.

A key first step was to know the Yang-Mills beta function. (I assumed [correctly] in my live talk that its definition had been made clear in the earlier remarks of my co-recipients; it is, after all, the minus sign to which I first alluded.) By the way, Erick Weinberg was supposed to compute it for an appendix of his thesis, to carry out a generalization of a renormalization group flow argument that appears in the Coleman-Weinberg paper, except for a realistic, non-Abelian weak interaction theory. But, in the end, I guess he figured he had enough stuff to get his degree, and it was time for him to move on to something new. I had actually hoped we'd compare notes, but he never attempted the calculation.

I visited Coleman a couple of times in Princeton. When I described to him my new, specific research program, I asked if he knew whether the beta function had already been computed. He thought not but said we should ask David Gross, who was down the hall. David said no, and we discussed briefly then that, while the calculation may have seemed to some to be daunting, it would, in fact, be straightforward.

Fortunately for both of us – and for Frank, too – he was probably wrong, though this episode is fraught with ambiguity. To my knowledge, there are no relevant printed records of the crucial bits of the story, which have been handed down only as folklore, existing in a variety of variant versions.

At a major particle physics meeting in Marseilles the previous year, attended by many particle physics luminaries, Symanzik gave a talk precisely about what came to be known as asymptotic freedom. He described how it could account for the otherwise mysterious results from SLAC on electron-proton scattering. Symanzik knew that the beta functions for other theories were all positive. In fact, many wise people thought there was a general, model-independent argument for positivity. For example, Schwinger[N] later asked me after hearing me speak on the subject, "What about the positivity of the spectral function?" I.e., intermediate, physical states come with positive probabilities. (This refers to an argument that is, indeed, relevant to other theories.) Symanzik said it would be interesting to know the answer for Yang-Mills theory, and then 't Hooft[N] announced it was negative. In some versions of the story, 't Hooft[N] spoke up at the question and answer period following Symanzik's talk. However, there are attendees of the meeting who have no such recollection. In other versions, it was a private exchange between 't Hooft[N] and Symanzik.

There are a variety of first-, second-, and third-hand accounts of why nothing further was heard on this subject from 't Hooft[N]. I won't repeat them here.* But I'd like to speculate a complementary perspective as to why no

* I will add one conjecture to the list, though it is not something I ever confirmed with 't Hooft[N]. It is possible that at that time 't Hooft[N] knew the sign of the beta function but not its coefficient. His calculations employed dimensional regularization and dimensional subtraction. From these he would have known the sign of the renormalization constants. However, the fundamental definition of the beta function makes reference to the response of the theory to scale transformations. Dimensional regularization introduces a scale in a subtle way – when one analytically continues away from the superficially scale invariant dimensions. How the traditional renormalization group is represented in this context is something that was worked out only a couple of years later.

one else at the meeting got wind of it or otherwise took any notice. (Admittedly, I wasn't there. So this is pure speculation.) Most theorists' attention then was on weak interactions, and this is a strong interaction issue. But that's not a good enough excuse. People did, after all, talk a lot about the scaling of the Stanford electron-proton experiments. Rather, I think Symanzik's speaking style played a crucial role. He was a charming, intense, sweet, and brilliant man. But his live delivery left something to be desired. I remember a different talk I heard him give on a somewhat related subject. He used hand-written slides for an overhead projector (which were the industry standard at the time for technical presentations). However, he obviously wrote out his slides with lined paper underneath as a guide, using every line. So he ended up with over twenty five lines of equations and text per page. His handwriting was typical German: undecipherable, at least to Americans, looking like endless up-down-up-down-up-down. The clincher, though, was when an equation on one page referred to an equation on another. He'd slap the second slide on top of the first, off-set the two by half a line, and point to both.

I slowly and carefully completed a calculation of the Yang-Mills beta function. I happen to be ambidextrous and mildly dyslexic. So I have trouble with left/right, in/out, forward/backward, etc. Hence, I derived each partial result from scratch, paying special attention to signs and conventions. It did not take long to go from dismay over the final minus sign (it was indeed useless for studying low energy phenomena) to excitement over the possibilities. I phoned Sidney Coleman. He listened patiently and said it was interesting. But, according to Coleman, I had apparently made an error because David Gross and his student had completed the same calculation, and they found it was plus. Coleman seemed to have more faith in the reliability of a team of two, which included a seasoned theorist, than in a single, young student. I said I'd check it yet once more. I called again about a week later to say I could find nothing wrong with my first calculation. Coleman said yes, he knew because the Princeton team had found a mistake, corrected it, and already submitted a paper to *Physical Review Letters.*

On learning of the Gross-Wilczek-Politzer result, Ken Wilson[N], who might have thought of its impossibility along the same lines as I attributed to Schwinger[N], above, knew who to call to check the result. He realized that there were actually several people around the world who had done the calculation, *en passant* as it were, as part of their work on radiative corrections to weak interactions in the newly-popular Weinberg-Salam model. They just never thought to focus particularly on this aspect. But they could quickly confirm for Wilson[N] by looking in their notebooks that the claimed result was, indeed, correct.

Steve Weinberg[N] and Murray Gell-Mann[N] were among those to instantly embrace non-Abelian color SU(3) gauge theory as the theory of the strong interactions. In Gell-Mann's case, it was in no small part because he had already invented it (!) with Harald Fritzsch and christened it QCD. He had previously articulated three solid arguments for choosing this particular theory. (For the physicists, those arguments were: baryon statistics, $\pi \to 2\gamma$, and the electron-positron annihilation cross section). And asymptotic freedom, i.e., the negative beta function, was the clincher. I'd only heard of Gell-Mann[N] and Fritzsch's work second hand, from Shelly Glashow[N], and he seemed think it shouldn't be taken too seriously. I only later realized it was more Glashow's mode of communication than his serious assessment of the plausibility of the proposal. In any case, I had completely lost track of Gell-Mann[N] and Fritzsch's QCD.

After the first seminar I ever gave on this subject (it was at MIT), I was approached by Ken Johnson (who, himself had done pioneering work on the renormalization group years earlier) and Vicki Weisskopf. "Very nice," they said. "Too bad that it is in glaring contradiction to at least two important classes of experiments." One problem was the electron-positron cross section, which had only gotten much worse since Gell-Mann[N] and Fritzsch's proposal of QCD, and the other was the issue of large angle products in proton-proton collisions. There were many more energetic particles produced than expected (naively) from QCD. By the way, this second issue attracted Dick Feynman's attention. And it wasn't until a couple of years later and his careful analysis with Rick Field that QCD was reconciled with those experiments. Only then did Feynman[N] join the ranks of the believers.

The experimentally measured electron-positron cross section (as a function of increasing collision energy) had leveled off – instead of continuing to drop steeply, which was thought to be a QCD prediction. In Aspen, Colorado, in the summer of 1974, I crossed paths with Ken Wilson[N], who, characteristically succinct, said, "It's charm, and it's not short distance." Tom Appelquist and I made it our task to understand those oracular comments and flesh out their consequences. By the end of the summer, the reconciliation of QCD with the experimental measurements was pretty clear to us. Tom toured the country explaining our work. His seminars included a sketch of what the cross section really was as opposed to what the experimentalists reported and an estimate, albeit technically an upper bound, on the astoundingly long lifetime of a particle that was being produced and decaying as yet unnoticed. Many people heard those talks and remember them, and there is at least one, objective written record of their existence: Sid Drell gave an account in a piece he wrote subsequently for *Scientific American* about charm. At the time, there were already many what-proved-to-be wrong papers trying to interpret the electron-positron experiments, and the SLAC experiment leader, Burt Richter[N], was touring the country explaining that he had made the monumental discovery that the electron was actually a little hadron, i.e., a strongly interacting particle like the proton, only much smaller in diameter. (This discovery, or at least the same experimental results, had been observed a few years earlier at the Cambridge Electron Accelerator, a joint Harvard-MIT venture. But no one believed it, and the machine was decommissioned.) Appelquist and I were drafting a paper. But I was the conservative one, perhaps overly influenced, I later realized, by a talk that I had heard by Steve Adler as to how large the discrepancy between naive QCD calculation and experimental measurements could be before the theory was in definite trouble. I focused on the things we could most reliably compute and did not appreciate the correctness of Tom's more general arguments.

In November that Fall came the experimental announcements. SLAC observed a particle (they called it the ψ) and ultimately observed a whole cross section just as predicted by Appelquist. And observation was simultaneously announced by Sam Ting$^{\mathcal{N}}$, in an experiment that identified a pimple, which Ting$^{\mathcal{N}}$ eponymously titled the $\mathcal{J}$[*], on what had been known as Lederman's Shoulder. (That's Leon Lederman$^{\mathcal{N}}$.) That is to say that Ting's experiment had actually been done earlier by Lederman$^{\mathcal{N}}$. The earlier experiment had cruder resolution, but it clearly indicated that there was something anomalous at just that energy.

Appelquist and I hurriedly dashed off a short version of our work to *Physical Review Letters*, where it was immediately and unequivocally rejected by senior editor Sy Pasternak. It was against that journal's policy to let authors engage in the coining of frivolous, new terminology. In the case at hand, our friend and colleague Alvaro De Rújula, on hearing of our work, had coined the term "charmonium", which in a single word was able to transmit the central new idea of the paper to any serious particle physics reader. Ultimately, Shelly Glashow$^{\mathcal{N}}$ brokered a compromise with Pasternak. We could use "charmonium" in the text but not in the title. The negotiations caused a delay of a couple of weeks – a long time in those heady days. As a consequence, publication came along side several other long-since-forgotten papers, instead of being hard on the heals of the experimental discovery.

That our explanation was correct was soon widely appreciated, and it convinced almost all of the remaining skeptics of the validity of QCD. I suspect that the consensus on this issue was a major contributing factor to the Royal Swedish Academy of Sciences' recognition within just a couple of years of Richter$^{\mathcal{N}}$ and Ting's discovery.

I hope you all now understand why I owe Tom Appelquist a huge, profound, and public apology. We certainly could have submitted for publication in September substantially the same paper we ultimately wrote two months later.

Now, somewhat out of chronological order, I'd like to express my thanks to my old friend and collaborator Howard Georgi. After the calculation of the beta function, it was fairly obvious what should be done next. One had to redo some calculations that had been done earlier by Norman Christ, Brosl Hasslacher, and Al Mueller but in the context of what was now, obviously, the right theory. Here, again, a missing name from that collaboration but who had a major impact was Georgio Parisi. Well, Howard Georgi checked up periodically on my progress, and I admitted having some technical trouble. So he volunteered to help, and we went on to do an enormous number of clever things together.

Apropos clever, there are some advances that require considerable mental struggle and lengthy argumentation, only to virtually disappear as non-issues,

*a reasonable approximation to the relevant Chinese character.

because they're simply obvious from a newer perspective. For example, the fact that quarks could have a mass, something unambiguously quantifiable and measured in grams – in spite of their never existing as isolated particles – was one such issue on which I battled with many physicists, including, for example, Gell-Mann[N] and Steve Weinberg[N]. The heavy charm quark gave impetus to those considerations, but there was a conceptual battle that had to be fought against older prejudices formed in the limited context of the "light" quarks. Younger physicists today can't even imagine that there was ever an issue.

Heavy quarks appeared once again in my research life. Joe Polchinski asked Mark Wise, a colleague of mine at Caltech, a question about heavy quark calculations, which Mark and I proceeded to answer. It was again a case where, unbeknownst to us, the work had already been done, this time by Misha Shifman and Mike Voloshin in the Soviet Union. Furthermore, I again missed the most important phenomenological consequences of that line of thought. Those had to wait for the collaboration of Mark Wise with Nathan Isgur. That heavy quark physics depends only trivially on the actual value of the heavy quark mass was obvious to me and probably most anyone else who gave it a thought. What Isgur and Wise noted was that in a world with more than one type of heavy quark, this gives rise to symmetries of monumentally useful importance. (The second heavy quark, the so-called bottom quark, was identified only several years after the first, i.e., the charmed quark.)

The establishment by the mid-1970's of QCD as the correct theory of the strong interactions completed what is now known prosaically as the Standard Model. It offers a description of all known fundamental physics except for gravity, and gravity is something that has no discernible effect when particles are studied a few at a time. However, the situation is a bit like the way that the Navier-Stokes equation accounts for the flow of water. The equations are at some level obviously correct, but there are only a few, limited circumstances in which their consequences can be worked out in any detail. Nevertheless, many leading physicists were inclined to conclude in the late 1970's that the task of basic physics was nearly complete, and we'd soon be out of jobs. A famous example was the inaugural lecture of Stephen Hawking as Lucasian Professor of Mathematics, a chair first held by Isaac Barrow at Cambridge University. Hawking titled his lecture, "Is the End in Sight for Theoretical Physics?" And he argued strongly for "Yes".

But more recent observations of astronomers have turned things on their heads. Recall, if you will, that among the many stupendous insights of Isaac Newton, the second Lucasian Professor, was the idea that the stuff of the heavens was the same stuff as matter here on Earth. This was revolutionary. And he asserted that the laws that governed the motion of stuff in the heavens were the same laws as applied to matter on Earth. (That there are laws at all may be his most profound insight. It is certainly what came to define the whole discipline of physics.) For three centuries we accumulated stunning detailed confirmation of these of Newton's assertions. But in a very fundamental way both of these ideas now appear to be about as wrong as they possibly could be

– at least that's the simplest interpretation of our current large-scale astrophysical observations. It turns out that we haven't a clue what virtually all of the matter in the universe consists of – except that it's not made of the particles that make up matter on Earth or in the stars. Furthermore, the force which governs the largest scale motions in the universe has nothing to do with the forces of the Standard Model or with gravity as it is familiar here on Earth.

There is a very active field of theoretical research which seeks to go beyond the Standard Model. Success in these endeavors would mean explaining the apparently arbitrary aspects of the Standard Model; success would mean bringing an account of gravity into the picture; and success would mean illuminating the previously mentioned issues in astrophysics. However, we now face a very serious problem in advancing the experimental frontier, a problem which few people like to discuss. It seems to me that ever since Leeuwenhoek, advances in the resolving power of our "microscopes" have come with similar investments of capital and manpower. I.e., an increase by an order of magnitude in the one required an increase by roughly an order of magnitude of the other – at least once we average over fits and starts and brilliant insights. The last big machine planned and canceled in the U.S. was to cost about \$10 billion. (That's $\$10^{10}$.) That would have allowed us to reach distances small enough to study the interactions of weak bosons directly. The realm of the conjectured "unification" of the forces of the Standard model, the realm of their possible unification with gravity, and the basic physics of String Theory, the most widely pursued approach to a physics more fundamental than the Standard Model, are all more than a dozen orders of magnitude further away. However, $\$10^{22}$ is simply not available for this line of research (or anything else for that matter).

The question of the benefit of this work incurred on mankind, an aspect stipulated in Mr. Nobel's will, is a whole other topic. But, as I said at the outset, I certainly appreciate the care and wisdom invested by the Royal Swedish Academy of Sciences in identifying noteworthy advances in fundamental physics – and in identifying the particular advance that we celebrate today. The reality of the actual progress of science is, however, often very complicated, as I hope I have conveyed from my few examples. The committees of the Academy know this full well, but their deliberations are confidential. I felt strongly that more of the public should contemplate these matters if they wish to understand not just the ideas of science but also how they have developed. I also hope that more of the scientific community would remember them, too.

My presentation in Stockholm ended at this point, but, in the days that followed, it prompted a variety of comments, questions, and exchanges. I'd like to add here a brief version of one of them. I was asked, point blank, what I actually thought of the 2004 Nobel Prizes in Physics, aside from the obvious personal considerations. And this is a distilled version of my reply. Recognition of the theory of the strong interactions is an obvious choice – for all the reasons that have been discussed in my co-recipients' lectures, in the presentation speech in Stockholm by Lars Brink, in the assembled material of the Nobel Foundation, and in the wide coverage elsewhere. However, in my view, getting to our current level of understanding has been a rich and complex story. Nevertheless, I believe that it is the overwhelming consensus (but by no means unanimous) opinion of researchers in the field – and I personally agree – that the discovery of asymptotic freedom was a genuinely crucial event. For some, it made everything clear. For others, it was only the beginning. And for yet others, it was the beginning of the final chapter. But in any case, it was key.

REFERENCES

The two books I would recommend first which give excellent accounts of this epoch in particle physics (and more) are:

1. *The Second Creation,* by Robert Crease and Charles Mann.
2. *Constructing Quarks,* by Andrew Pickering.

The first is colorful in its rendition of the personalities, rather accurate in its physics, and totally accessible to the interested layman. The second is a more scholarly endeavor. Pickering began his career as a particle theorist, a contemporary of mine. He includes considerable scientific detail but still aims at a non-technical audience.

Two marvelous books on twentieth century physics for the interested layman which focus more on scientific substance rather than historical process are:

3. *The Cosmic Code*
4. *Perfect Symmetry*

both by Heinz Pagels, something of a self-styled New York City dandy, but as charming a person as could be. He died young but just as he dreamed it would be.

The Cosmic Code is about quantum mechanics, and *Perfect Symmetry* covers more of the sweep of particle physics and cosmology.

I have not sought out the actual published references for the relevant points in my narrative. They're not hard to find, but this is not a refereed journal. And yet, there is a potentially enlightening aspect to my having put this together purely from memory in October and November of 2004. While standard references are unequivocally available in the published record, what actually transpired, leading to those publications, is not. We rely on people's personal accounts. And now we enter the interesting realm where participants in the same event may have very different and mutually contradictory perceptions of what transpired, *and* those perceptions may shift as time passes. While intentional deception is not an unheard of phenomenon, these phenomena effect the reports of people with the highest integrity. Although evaluating the accuracy of my personal recollections may be very difficult, at least it would be possible to see how good my memory is with respect to items that can be confirmed or refuted.

ASYMPTOTIC FREEDOM: FROM PARADOX TO PARADIGM

Nobel Lecture, December 8, 2004

by

FRANK A. WILCZEK

Massachusetts Institute of Technology (MIT), Cambridge, USA.

1 A PAIR OF PARADOXES

In theoretical physics, paradoxes are good. That's paradoxical, since a paradox appears to be a contradiction, and contradictions imply serious error. But Nature cannot realize contradictions. When our physical theories lead to paradox we must find a way out. Paradoxes focus our attention, and we think harder.

When David Gross and I began the work that led to this Nobel Prize [1, 2, 3, 4], in 1972, we were driven by paradoxes. In resolving the paradoxes we were led to discover a new dynamical principle, asymptotic freedom. This principle in turn has led to an expanded conception of fundamental particles, a new understanding of how matter gets its mass, a new and much clearer picture of the early universe, and new ideas about the unity of Nature's forces. Today I'd like to share with you the story of these ideas.

1.1 *Paradox 1: Quarks are Born Free, but Everywhere They are in Chains*

The first paradox was phenomenological.

Near the beginning of the twentieth century, after pioneering experiments by Rutherford, Geiger and Marsden, physicists discovered that most of the mass and all of the positive charge inside an atom is concentrated in a tiny central nucleus. In 1932, Chadwick discovered neutrons, which together with protons could be considered as the ingredients out of which atomic nuclei could be constructed. But the known forces, gravity and electromagnetism, were insufficient to bind protons and neutrons tightly together into objects as small as the observed nuclei. Physicists were confronted with a new force, the most powerful in Nature. It became a major challenge in fundamental physics, to understand this new force.

For many years physicists gathered data to address that challenge, basically by bashing protons and neutrons together and studying what came out. The results that emerged from these studies, however, were complicated and hard to interpret.

What you would expect, if the particles were really fundamental (indestructible), would be the same particles you started with, coming out with just their trajectories changed. Instead, the outcome of the collisions was often many particles. The final state might contain several copies of the originals, or different particles altogether. A plethora of new particles was discovered in this way. Although these particles, generically called hadrons, are unstable, they otherwise behave in ways that broadly resemble the way protons and neutrons behave. So the character of the subject changed. It was no longer natural to think of it as simply as the study of a new force that binds protons and neutrons into atomic nuclei. Rather, a new world of phenomena had come into view. This world contained many unexpected new particles, that could transform into one another in a bewildering variety of ways. Reflecting this change in perspective, there was a change in terminology. Instead of the nuclear force, physicists came to speak of the strong interaction.

In the early 1960s, Murray Gell-Mann and George Zweig made a great advance in the theory of the strong interaction, by proposing the concept of quarks. If you imagined that hadrons were not fundamental particles, but rather that they were assembled from a few more basic types, the quarks, patterns clicked into place. The dozens of observed hadrons could be understood, at least roughly, as different ways of putting together just three kinds ("flavors") of quarks. You can have a given set of quarks in different spatial orbits, or with their spins aligned in different ways. The energy of the configuration will depend on these things, and so there will be a number of states with different energies, giving rise to particles with different masses, according to $m = E/c^2$. It is analogous to the way we understand the spectrum of excited states of an atom, as arising from different orbits and spin alignments of electrons. (For electrons in atoms the interaction energies are relatively small, however, and the effect of these energies on the overall *mass* of the atoms is insignificant.)

The rules for using quarks to model reality seemed quite weird, however.

Quarks were supposed to hardly notice one another when they were close together, but if you tried to isolate one, you found that you could not. People looked very hard for individual quarks, but without success. Only bound states of a quark and an antiquark – mesons – or bound states of three quarks – baryons – are observed. This experimental regularity was elevated into The Principle of Confinement. But giving it a dignified name didn't make it less weird.

There were other peculiar things about quarks. They were supposed to have electric charges whose magnitudes are fractions ($\frac{2}{3}$ or $\frac{1}{3}$) of what appears to be the basic unit, namely the magnitude of charge carried by an electron or proton. All other observed electric charges are known, with great accuracy, to be whole-number multiples of this unit. Also, identical quarks did not appear to obey the normal rules of quantum statistics. These rules would require that, as spin $\frac{1}{2}$ particles, quarks should be fermions, with antisymmetric wave functions. The pattern of observed baryons cannot be understood using antisymmetric wave functions; it requires symmetric wave functions.

The atmosphere of weirdness and peculiarity surrounding quarks thickened into paradox when J. Friedman, H. Kendall, R. Taylor and their collaborators at the Stanford Linear Accelerator (SLAC) used energetic photons to poke into the inside of protons [5]. They discovered that there are indeed entities that look like quarks inside protons. Surprisingly, though, they found that when quarks are hit hard they seem to move (more accurately: to transport energy and momentum) as if they were free particles. Before the experiment, most physicists had expected that whatever caused the strong interaction of quarks would also cause quarks to radiate energy abundantly, and thus rapidly to dissipate their motion, when they got violently accelerated.

At a certain level of sophistication, that association of radiation with forces appears inevitable, and profound. Indeed, the connection between forces and radiation is associated with some of the most glorious episodes in the history of physics. In 1864, Maxwell predicted the existence of electromagnetic radiation – including, but not limited to, ordinary light – as a consequence of his consistent and comprehensive formulation of electric and magnetic forces. Maxwell's new radiation was subsequently generated and detected by Hertz, in 1883 (and over the twentieth century its development has revolutionized the way we manipulate matter and communicate with one another). Much later, in 1935, Yukawa predicted the existence of pions based on his analysis of nuclear forces, and they were subsequently discovered in the late 1940s; the existences of many other hadrons were predicted successfully using a generalization of these ideas. (For experts: I have in mind the many resonances that were first seen in partial wave analyses, and then later in production.) More recently the existence of *W* and *Z* bosons, and of color gluons, and their properties, was inferred before their experimental discovery. Those discoveries were, in 1972, still ahead of us, but they serve to confirm, retroactively, that our concerns were worthy ones. Powerful interactions ought to be associated with powerful radiation. When the most powerful interaction in nature, the strong interaction, did not obey this rule, it posed a sharp paradox.

1.2 *Paradox 2: Special Relativity and Quantum Mechanics Both Work*

The second paradox is more conceptual. Quantum mechanics and special relativity are two great theories of twentieth-century physics. Both are very successful. But these two theories are based on entirely different ideas, which are not easy to reconcile. In particular, special relativity puts space and time on the same footing, but quantum mechanics treats them very differently. This leads to a creative tension, whose resolution has led to three previous Nobel Prizes (and ours is another).

The first of these prizes went to P. A. M. Dirac (1933). Imagine a particle moving on average at very nearly the speed of light, but with an uncertainty in position, as required by quantum theory. Evidently it there will be some probability for observing this particle to move a little faster than average, and therefore faster than light, which special relativity won't permit. The only known way to resolve this tension involves introducing the idea of antiparticles. Very roughly speaking, the required uncertainty in position is accommodated by allowing for the possibility that the act of measurement can involve the creation of several particles, each indistinguishable from the original, with different positions. To maintain the balance of conserved quantum numbers, the extra particles must be accompanied by an equal number of antiparticles. (Dirac was led to predict the existence of antiparticles through a sequence of ingenious interpretations and re-interpretations of the elegant relativistic wave equation he invented, rather than by heuristic reasoning of the sort I've presented. The inevitability and generality of his conclusions, and their direct relationship to basic principles of quantum mechanics and special relativity, are only clear in retrospect).

The second and third of these prizes were to R. Feynman, J. Schwinger, and S.-I. Tomonaga (1965) and to G. 't Hooft and M. Veltman (1999) respectively. The main problem that all these authors in one way or another addressed is the problem of ultraviolet divergences.

When special relativity is taken into account, quantum theory must allow for fluctuations in energy over brief intervals of time. This is a generalization of the complementarity between momentum and position that is fundamental for ordinary, non-relativistic quantum mechanics. Loosely speaking, energy can be borrowed to make evanescent virtual particles, including particle-antiparticle pairs. Each pair passes away soon after it comes into being, but new pairs are constantly boiling up, to establish an equilibrium distribution. In this way the wave function of (superficially) empty space becomes densely populated with virtual particles, and empty space comes to behave as a dynamical medium.

The virtual particles with very high energy create special problems. If you calculate how much the properties of real particles and their interactions are changed by their interaction with virtual particles, you tend to get divergent answers, due to the contributions from virtual particles of very high energy.

This problem is a direct descendant of the problem that triggered the introduction of quantum theory in the first place, i.e. the "ultraviolet catastrophe" of black body radiation theory, addressed by Planck. There the problem was that high-energy modes of the electromagnetic field are predicted, classically, to occur as thermal fluctuations, to such an extent that equilibrium at any finite temperature requires that there is an infinite amount of energy in these modes. The difficulty came from the possibility of small-amplitude fluctuations with rapid variations in space and time. The element of discreteness introduced by quantum theory eliminates the possibility of very small-amplitude fluctuations, because it imposes a lower bound on their size. The (relatively) large-amplitude fluctuations that remain are predicted to occur very rarely in thermal equilibrium, and cause no problem. But quantum fluctuations are much more efficient than are thermal fluctuations at exciting the high-energy modes, in the form of virtual particles, and so those modes come back to haunt us. For example, they give a divergent contribution to the energy of empty space, the so-called zero-point energy.

Renormalization theory was developed to deal with this sort of difficulty. The central observation that is exploited in renormalization theory is that although interactions with high-energy virtual particles appear to produce divergent corrections, they do so in a very structured way. That is, the same corrections appear over and over again in the calculations of many different physical processes. For example in quantum electrodynamics (QED) exactly two independent divergent expressions appear, one of which occurs when we calculate the correction to the mass of the electron, the other of which occurs when we calculate the correction to its charge. To make the calculation mathematically well-defined, we must artificially exclude the highest energy modes, or dampen their interactions, a procedure called applying a cut-off, or regularization. In the end we want to remove the cut-off, but at intermediate stages we need to leave it in, so as to have well-defined (finite) mathematical expressions. If we are willing to take the mass and charge of the electron from experiment, we can identify the formal expressions for these quantities, including the potentially divergent corrections, with their measured values. Having made this identification, we can remove the cutoff. We thereby obtain well-defined answers, in terms of the measured mass and charge, for everything else of interest in QED.

Feynman, Schwinger, and Tomonoga developed the technique for writing down the corrections due to interactions with any finite number of virtual particles in QED, and showed that renormalization theory worked in the simplest cases. (I'm being a little sloppy in my terminology; instead of saying the number of virtual particles, it would be more proper to speak of the number of internal loops in a Feynman graph.) Freeman Dyson supplied a general proof. This was intricate work, that required new mathematical techniques. 't Hooft and Veltman showed that renormalization theory applied to a much wider class of theories, including the sort of spontaneously broken gauge theories that had been used by Glashow, Salam, and Weinberg to construct the (now) standard model of electroweak interactions. Again, this was intricate and highly innovative work.

This brilliant work, however, still did not eliminate all the difficulties. A very profound problem was identified by Landau [6]. Landau argued that virtual particles would tend to accumulate around a real particle as long as there was any uncancelled influence. This is called screening. The only way for this screening process to terminate is for the source plus its cloud of virtual particles to cease to be of interest to additional virtual particles. But then, in the end, no uncancelled influence would remain – and no interaction!

Thus all the brilliant work in QED and more general field theories represented, according to Landau, no more than a temporary fix. You could get finite results for the effect of any particular number of virtual particles, but when you tried to sum the whole thing up, to allow for the possibility of an arbitrary number of virtual particles, you would get nonsense – either infinite answers, or no interaction at all.

Landau and his school backed up this intuition with calculations in many different quantum field theories. They showed, in all the cases they calculated, that screening in fact occurred, and that it doomed any straightforward attempt to perform a complete, consistent calculation by adding up the contributions of more and more virtual particles. We can sweep this problem under the rug in QED or in electroweak theory, because the answers including only a small finite number of virtual particles provide an excellent fit to experiment, and we make a virtue of necessity by stopping there. But for the strong interaction that pragmatic approach seemed highly questionable, because there is no reason to expect that lots of virtual particles won't come into play, when they interact strongly.

Landau thought that he had destroyed quantum field theory as a way of reconciling quantum mechanics and special relativity. Something would have to give. Either quantum mechanics or special relativity might ultimately fail, or else essentially new methods would have to be invented, beyond quantum field theory, to reconcile them. Landau was not displeased with this conclusion, because in practice quantum field theory had not been very helpful in understanding the strong interaction, even though a lot of effort had been put into it. But neither he, nor anyone else, proposed a useful alternative.

So we had the paradox, that combining quantum mechanics and special relativity seemed to lead inevitably to quantum field theory; but quantum field theory, despite substantial pragmatic success, self-destructed logically due to catastrophic screening.

2 PARADOX LOST: ANTISCREENING, OR ASYMPTOTIC FREEDOM

These paradoxes were resolved by our discovery of asymptotic freedom.

We found that some very special quantum field theories actually have antiscreening. We called this property asymptotic freedom, for reasons that will soon be clear. Before describing the specifics of the theories, I'd like to indicate in a rough, general way how the phenomenon of antiscreening allows us to resolve our paradoxes.

Antiscreening turns Landau's problem on its head. In the case of screening, a source of influence – let us call it charge, understanding that it can represent something quite different from electric charge – induces a canceling cloud of virtual particles. From a large charge, at the center, you get a small observable influence far away. Antiscreening, or asymptotic freedom, implies instead that a charge of intrinsically small magnitude catalyzes a cloud of virtual particles that enhances its power. I like to think of it as a thundercloud that grows thicker and thicker as you move away from the source.

Since the virtual particles themselves carry charge, this growth is a self-reinforcing, runaway process. The situation appears to be out of control. In particular, energy is required to build up the thundercloud, and the required energy threatens to diverge to infinity. If that is the case, then the source could never be produced in the first place. We've discovered a way to avoid Landau's disease – by banishing the patients!

At this point our first paradox, the confinement of quarks, makes a virtue of theoretical necessity. For it suggests that there *are* in fact sources – specifically, quarks – that cannot exist on their own. Nevertheless, Nature teaches us, these confined particles can play a role as building-blocks. If we have, nearby to a source particle, its antiparticle (for example, quark and antiquark), then the catastrophic growth of the antiscreening thundercloud is no longer inevitable. For where they overlap, the cloud of the source can be canceled by the anticloud of the antisource. Quarks and antiquarks, bound together, can be accommodated with finite energy, though either in isolation would cause an infinite disturbance.

Because it was closely tied to detailed, quantitative experiments, the sharpest problem we needed to address was the paradoxical failure of quarks to radiate when Friedman, Kendall, and Taylor subjected them to violent acceleration. This too can be understood from the physics of antiscreening. According to this mechanism, the color charge of a quark, viewed up close, is small. It builds up its power to drive the strong interaction by accumulating a growing cloud at larger distances. Since the power of its intrinsic color charge is small, the quark is actually only loosely attached to its cloud. We can jerk it away from its cloud, and it will – for a short while – behave almost as if it had no color charge, and no strong interaction. As the virtual particles in space respond to the altered situation they rebuild a new cloud, moving along with the quark, but this process does not involve significant radiation of energy and momentum. That, according to us, was why you could analyze the most salient aspects of the SLAC experiments – the inclusive cross-sections, which only keep track of overall energy-momentum flow – as if the quarks were free particles, though in fact they are strongly interacting and ultimately confined.

Thus both our paradoxes, nicely dovetailed, get resolved together through antiscreening.

The theories that we found to display asymptotic freedom are called nonabelian gauge theories, or Yang-Mills theories [7]. They form a vast generalization of electrodynamics. They postulate the existence of several different kinds of charge, with complete symmetry among them. So instead of one entity, "charge", we have several "colors". Also, instead of one photon, we have a family of color gluons.

The color gluons themselves carry color charges. In this respect the nonabelian theories differ from electrodynamics, where the photon is electrically neutral. Thus gluons in nonabelian theories play a much more active role in the dynamics of these theories than do photons in electrodynamics. Indeed, it is the effect of virtual gluons that is responsible for antiscreening, which does not occur in QED.

It became evident to us very early on that one particular asymptotically free theory was uniquely suited as a candidate to provide the theory of the strong interaction. On phenomenological grounds, we wanted to have the possibility to accommodate baryons, based on three quarks, as well as mesons, based on quark and antiquark. In light of the preceding discussion, this requires that the color charges of three different quarks can cancel, when you add them up. That can oocur if the three colors exhaust all possibilities; so we arrived at the gauge group $SU(3)$, with three colors, and eight gluons. To be fair, several physicists had, with various motivations, suggested the existence of a three-valued internal color label for quarks years before [8]. It did not require a great leap of imagination to see how we could adapt those ideas to our tight requirements.

By using elaborate technical machinery of quantum field theory (including the renormalization group, operator product expansions, and appropriate dispersion relations) we were able to be much more specific and quantitative about the implications our theory than my loose pictorial language suggests. In particular, the strong interaction does not simply turn off abruptly, and there is a non-zero probability that quarks will radiate when poked. It is only asymptotically, as energies involved go to infinity, that the probability for radiation vanishes. We could calculate in great detail the observable effects of the radiation at finite energy, and make experimental predictions based on these calculations. At the time, and for several years later, the data was not accurate enough to test these particular predictions, but by the end of the 1970s they began to look good, and by now they're beautiful.

Our discovery of asymptotic freedom, and its essentially unique realization in quantum field theory, led us to a new attitude towards the problem of the strong interaction. In place of the broad research programs and fragmentary insights that had characterized earlier work, we now had a single, specific candidate theory – a theory that could be tested, and perhaps falsified, but which could not be fudged. Even now, when I re-read our declaration [3]

> Finally let us recall that the proposed theories appear to be uniquely singled out by nature, if one takes both the SLAC results and the renormalization-group approach to quantum field theory at face value.

I re-live the mixture of exhilaration and anxiety that I felt at the time.

3 A FOURSOME OF PARADIGMS

Our resolution of the paradoxes that drove us had ramifications in unanticipated directions, and extending far beyond their initial scope.

3.1 *Paradigm 1: The Hard Reality of Quarks and Gluons*

Because, in order to fit the facts, you had to ascribe several bizarre properties to quarks – paradoxical dynamics, peculiar charge, and anomalous statistics – their "reality" was, in 1972, still very much in question. This despite the fact that they were helpful in organizing the hadrons, and even though Friedman, Kendall, and Taylor had "observed" them! The experimental facts wouldn't go away, of course, but their ultimate significance remained doubtful. Were quarks basic particles, with simple properties, that could be used to in formulating a profound theory – or just a curious intermediate device, that would need to be replaced by deeper conceptions?

Now we know how the story played out, and it requires an act of imagination to conceive how it might have been different. But Nature is imaginative, as are theoretical physicists, and so it's not impossible to fantasize alternative

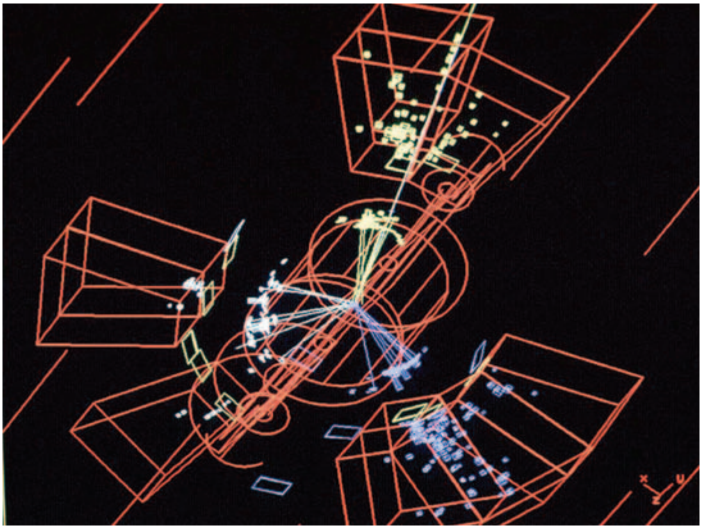


Figure 1: A photograph from the L3 collaboration, showing three jets emerging from electron-positron annihilation at high energy [9]. These jets are the materialization of a quark, antiquark, and gluon.

histories. For example, the quasiparticles of the fractional quantum Hall effect, which are not basic but rather emerge as collective excitations involving ordinary electrons, also cannot exist in isolation, and they have fractional charge and anomalous statistics! Related things happen in the Skyrme model, where nucleons emerge as collective excitations of pions. One might have fantasized that quarks would follow a similar script, emerging somehow as collective excitations of hadrons, or of more fundamental preons, or of strings.

Together with the new attitude toward the strong interaction problem, that I just mentioned, came a new attitude toward quarks and gluons. These words were no longer just names attached to empirical patterns, or to notional building blocks within rough phenomenological models. Quarks and (especially) gluons had become ideally simple entities, whose properties are fully defined by mathematically precise algorithms.

You can even see them! Here's a picture, which I'll now explain.

Asymptotic freedom is a great boon for experimental physics, because it leads to the beautiful phenomenon of jets. As I remarked before, an important part of the atmosphere of mystery surrounding quarks arose from the fact that they could not be isolated. But if we change our focus, to follow flows of energy and momentum rather than individual hadrons, then quarks and gluons come into view, as I'll now explain.

There is a contrast between two different kinds of radiation, which expresses the essence of asymptotic freedom. Hard radiation, capable of significantly re-directing the flow of energy and momentum, is rare. But soft radiation, that produces additional particles moving in the same direction, without deflecting the overall flow, is common. Indeed, soft radiation is associated

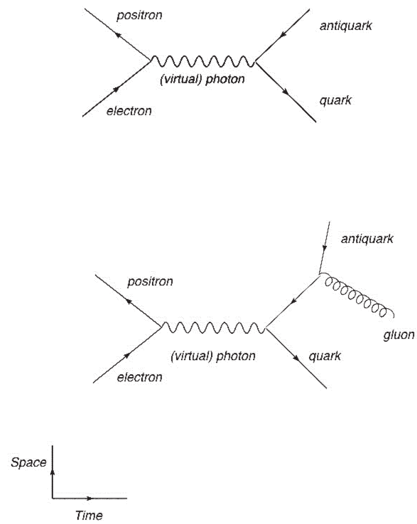


Figure 2: These Feynman graphs are schematic representations of the fundamental processes in electron-positron annihilation, as they take place in space and time. They show the origin of two-jet and three-jet events.

with the build-up of the clouds I discussed before, as it occurs in time. Let's consider what it means for experiments, say to be concrete the sort of experiment done at the Large Electron Positron collider (LEP) at CERN during the 1990s, and contemplated for the International Linear Collider (ILC) in the future. At these facilities, one studies what emerges from the annihilation of electrons and positrons that collide at high energies. By well-understood processes that belong to QED or electroweak theory, the annihilation proceeds through a virtual photon or *Z* boson into a quark and an antiquark. Conservation and energy and momentum dictate that the quark and antiquark will be moving at high speed in opposite directions. If there is no hard radiation, then the effect of soft radiation will be to convert the quark into a spray of hadrons moving in a common direction: a jet. Similarly, the antiquark becomes a jet moving in the opposite direction. The observed result is then a 2-jet event. Occasionally (about 10% of the time, at LEP) there will be hard radiation, with the quark (or antiquark) emitting a gluon in a significantly new direction. From that point on the same logic applies, and we have a 3-jet event, like the one shown in Figure 1. The theory of the underlying space-time process is depicted in Figure 2. And roughly 1% of the time 4 jets will occur, and so forth. The relative probability of different numbers of jets, how it varies with the overall energy, the relative frequency of different angles at which the jets emerge and the total energy in each – all these detailed aspects of the "antenna pattern" can be predicted quantitatively. These predictions reflect the basic couplings among quarks and gluons, which define QCD, quite directly.

The predictions agree well with very comprehensive experimental measurements. So we can conclude with confidence that QCD is right, and that what

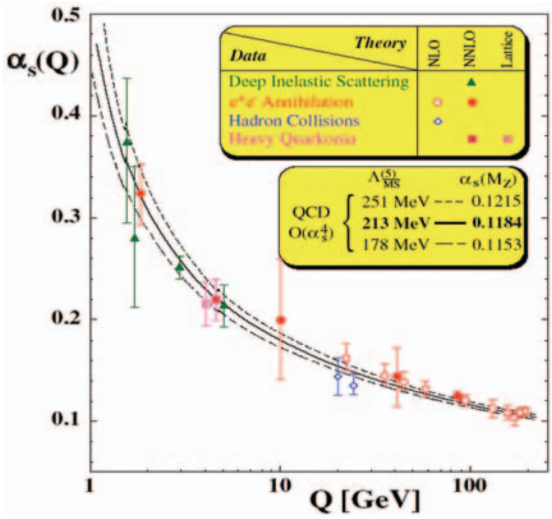


Figure 3: Many quite different experiments, performed at different energies, have been successfully analyzed using QCD. Each fits a large quantity of data to a single parameter, the strong coupling α_s. By comparing the values they report, we obtain direct confirmation that the coupling evolves as predicted [10].

you are seeing, in Figure 1, is a quark, an antiquark, and a gluon – although, since the predictions are statistical, we can't say for sure which is which!

By exploiting the idea that hard radiation processes, reflecting fundamental quark and gluon interactions, control the overall flow of energy and momentum in high-energy processes, one can analyze and predict the behavior of many different kinds of experiments. In most of these applications, including the original one to deep inelastic scattering, the analysis necessary to separate out hard and soft radiation is much more involved and harder to visualize than in the case of electron-positron annihilation. A lot of ingenuity has gone, and continues to go, into this subject, known as perturbative QCD. The results have been quite successful and gratifying. Figure 3 shows one aspect of the success. Many different kinds of experiments, performed at many different energies, have been successfully described by QCD predictions, each in terms of the one relevant parameter of the theory, the overall coupling strength. Not only must each experiment, which may involve hundreds of independent measurements, be fit consistently, but one can then check whether the values of the coupling change with the energy scale in the way we predicted. As you can see, it does. A remarkable tribute to the success of the theory, which I've been amused to watch evolve, is that a lot of the same activity that used to be called *testing* QCD is now called *calculating backgrounds*.

As a result of all this success, a new paradigm has emerged for the operational meaning of the concept of a fundamental particle. Physicists designing and interpreting high-energy experiments now routinely describe their results in terms of producing and detecting quarks and gluons: what they mean, of course, is the corresponding jets.

3.2 *Paradigm 2: Mass Comes from Energy*

My friend and mentor Sam Treiman liked to relate his experience of how, during World War II, the U.S. Army responded to the challenge of training a large number of radio engineers starting with very different levels of preparation, ranging down to near zero. They designed a crash course for it, which Sam took. In the training manual, the first chapter was devoted to Ohm's three laws. Ohm's first law is $V = IR$. Ohm's second law is $I = V/R$. I'll leave it to you to reconstruct Ohm's third law.

Similarly, as a companion to Einstein's famous equation $E = mc^2$ we have his second law, $m = E/c^2$.

All this isn't quite as silly as it may seem, because different forms of the same equation can suggest very different things. The usual way of writing the equation, $E = mc^2$, suggests the possibility of obtaining large amounts of energy by converting small amounts of mass. It brings to mind the possibilities of nuclear reactors, or bombs. Stated as $m = E/c^2$, Einstein's law suggests the possibility of explaining mass in terms of energy. That is a good thing to do, because in modern physics energy is a more basic concept than mass. Actually, Einstein's original paper does not contain the equation $E = mc^2$, but rather $m = E/c^2$. In fact, the title is a question: "Does the Inertia of a Body Depend Upon its Energy Content?" From the beginning, Einstein was thinking about the origin of mass, not about making bombs.

Modern QCD answers Einstein's question with a resounding "Yes!" Indeed, the mass of ordinary matter derives almost entirely from energy – the energy of massless gluons and nearly massless quarks, which are the ingredients from which protons, neutrons, and atomic nuclei are made.

The runaway build-up of antiscreening clouds, which I described before, cannot continue indefinitely. The resulting color fields would carry infinite energy, which is not available. The color charge that threatens to induce this runaway must be cancelled. The color charge of a quark can be cancelled either with an antiquark of the opposite color (making a meson), or with two quarks of the complementary colors (making a baryon). In either case, perfect cancellation would occur only if the particles doing the canceling were located right on top of the original quark – then there would be no uncanceled source of color charge anywhere in space, and hence no color field. Quantum mechanics does not permit this perfect cancellation, however. The quarks and antiquarks are described by wave functions, and spatial gradients in these wave function cost energy, and so there is a high price to pay for localizing the wave function within a small region of space. Thus, in seeking to minimize the energy, there are two conflicting considerations: to minimize the field energy, you want to cancel the sources accurately; but to minimize the wave-

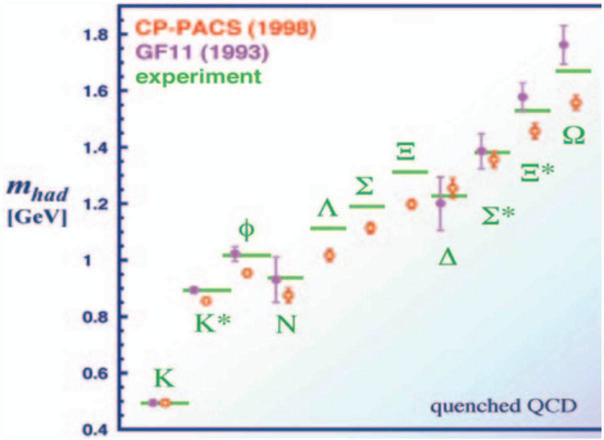


Figure 4: Comparison of observed hadron masses to the energy spectrum predicted by QCD, upon direct numerical integration of the equations, exploiting immense computer power [11]. The small remaining discrepancies are consistent with what is expected given the approximations that were necessary to make the calculation practical.

function localization energy, you want to keep the sources fuzzy. The stable configurations will be based on different ways of compromising between those two considerations. In each such configuration, there will be both field energy and localization energy. This gives rise to mass, according to $m = E/c^2$, even if the gluons and quarks started out without any non-zero mass of their own. So the different stable compromises will be associated with particles that we can observe, with different masses; and metastable compromises will be associated with observable particles that have finite lifetimes.

To determine the stable compromises concretely, and so to predict the masses of mesons and baryons, is hard work. It requires difficult calculations that continue to push the frontiers of massively parallel processing. I find it quite ironical that if we want to compute the mass of a proton, we need to deploy something like 10^{30} protons and neutrons, doing trillions of multiplications per second, working for months, to do what one proton does in 10^{-24} seconds, namely figure out its mass. Maybe it qualifies as a paradox. At the least, it suggests that there may be much more efficient ways to calculate than the ones we're using.

In any case, the results that emerge from these calculations are very gratifying. They are displayed in Figure 4. The observed masses of prominent mesons and baryons are reproduced quite well, stating from an extremely tight and rigid theory. Now is the time to notice also that one of the data points in Figure 3, the one labeled "Lattice", is of a quite different character from the others. It is based not on the perturbative physics of hard radiation, but rather on the comparison of a direct integration of the full equations of QCD with experiment, using the techniques of lattice gauge theory.

The success of these calculations represents the ultimate triumph over our two paradoxes:

- The calculated spectrum does not contain anything with the charges or other quantum numbers of quarks; nor of course does it contain massless gluons. The observed particles do not map in a straightforward way to the primary fields from which they ultimately arise.
- Lattice discretization of the quantum field theory provides a cutoff procedure that is independent of any expansion in the number of virtual particle loops. The renormalization procedure must be, and is, carried out without reference to perturbation theory, as one takes the lattice spacing to zero. Asymptotic freedom is crucial for this, as I discussed – it saves us from Landau's catastrophe.

By fitting some fine details of the pattern of masses, one can get an estimate of what the quark masses are, and how much their masses are contributing to the mass of the proton and neutron. It turns out that what I call QCD Lite – the version in which you put the u and d quark masses to zero, and ignore the other quarks entirely – provides a remarkably good approximation to reality. Since QCD Lite is a theory whose basic building-blocks have zero mass, this result quantifies and makes precise the idea that most of the mass of ordinary matter – 90 % or more – arises from pure energy, via $m = E/c^2$.

The calculations make beautiful images, if we work to put them in eye-friendly form. Derek Leinweber has made some striking animations of QCD fields as they fluctuate in empty space. Figure 5 is a snapshot from one of his animations. Figure 6 from Greg Kilcup, displays the (average) color fields, over and above the fluctuations, that are associated with a very simple hadron, the pion, moving through space-time. Insertion of a quark-antiquark pair, which we subsequently remove, produces this disturbance in the fields.

These pictures make it clear and tangible that the quantum vacuum is a dynamic medium, whose properties and responses largely determine the behavior of matter. In quantum mechanics, energies are associated with frequencies, according to the Planck relation $E = h\nu$. The masses of hadrons, then, are uniquely associated to tones emitted by the dynamic medium of space when it is disturbed in various ways, according to

$$\nu = mc^2/h \tag{1}$$

We thereby discover, in the reality of masses, an algorithmic, precise Music of the Void. It is a modern embodiment of the ancients' elusive, mystical "Music of the Spheres".

3.3 *Paradigm 3: The Early Universe was Simple*

In 1972 the early universe seemed hopelessly opaque. In conditions of ultra-high temperatures, as occurred close to the Big Bang singularity, one would have lots of hadrons and antihadrons, each one an extended entity that inter-

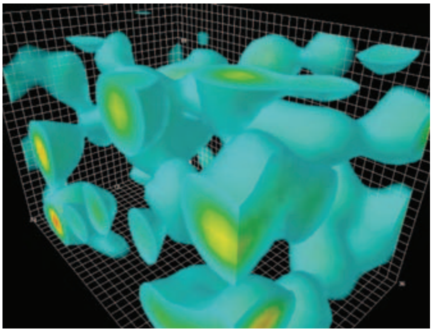


Figure 5: A snapshot of spontaneous quantum fluctuations in the gluon fields [12]. For experts: what is shown is the topological charge density in a typical contribution to the functional integral, with high frequency modes filtered out.

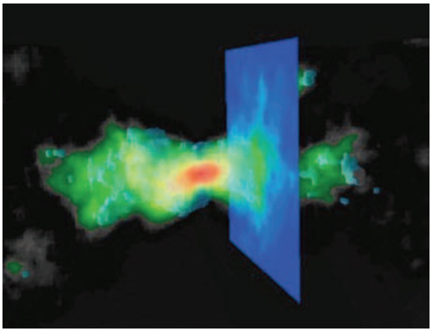


Figure 6: The calculated net distribution of field energy caused by injecting and removing a quark-antiquark pair [13]. By calculating the energy in these fields, and the energy in analogous fields produced by other disturbances, we predict the masses of hadrons. In a profound sense, these fields are the hadrons.

acts strongly and in complicated ways with its neighbors. They'd start to overlap with one another, and thereby produce a theoretically intractable mess.

But asymptotic freedom renders ultra-high temperatures friendly to theorists. It says that if we switch from a description based on hadrons to a description based on quark and gluon variables, and focus on quantities like total energy, that are not sensitive to soft radiation, then the treatment of the strong interaction, which was the great difficulty, becomes simple. We can calculate to a first approximate by pretending that the quarks, antiquarks and gluons behave as free particles, then add in the effects of rare hard interactions. This makes it quite practical to formulate a precise description of the properties of ultra-high temperature matter that are relevant to cosmology.

We can even, over an extremely limited volume of space and time, reproduce Big Bang conditions in terrestrial laboratories. When heavy ions are caused to collide at high energy, they produce a fireball that briefly attains temperatures as high as 200 MeV. "Simple" may not be the word that occurs to you in describing the explosive outcome of this event, as displayed in Figure 7, but in fact detailed study does permit us to reconstruct aspects of the initial fireball, and to check that it was a plasma of quarks and gluons.

3.4 *Paradigm 4: Symmetry Rules*

Over the course of the twentieth century, symmetry has been immensely fruitful as a source of insight into Nature's basic operating principles. QCD, in particular, is constructed as the unique embodiment of a huge symmetry group, local $SU(3)$ color gauge symmetry (working together with special relativity, in the

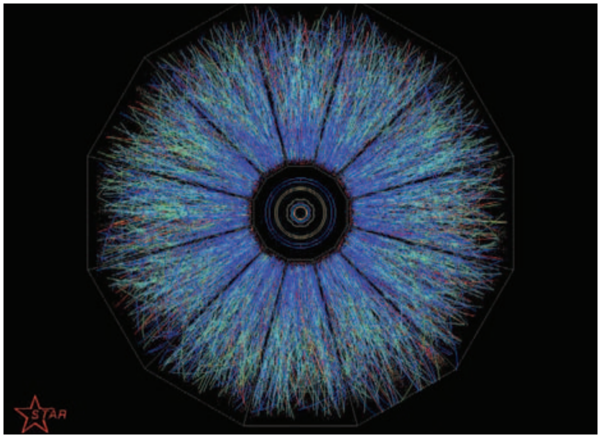


Figure 7: A picture of particle tracks emerging from the collision of two gold ions at high energy. The resulting fireball and its subsequent expansion recreate, on a small scale and briefly, physical conditions that last occurred during the Big Bang [14].

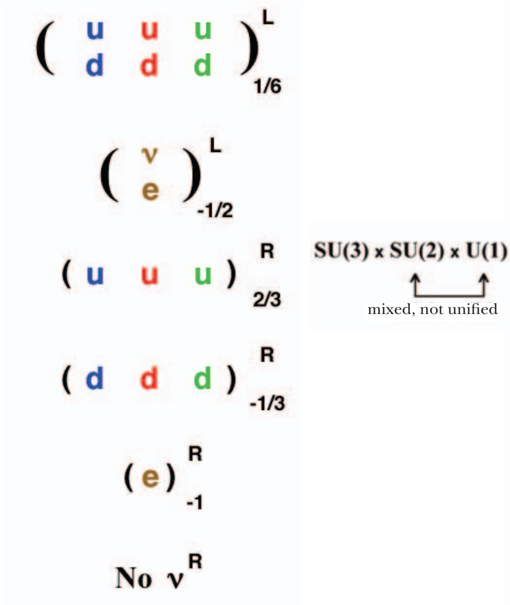
context of quantum field theory). As we try to discover new laws, that improve on what we know, it seems good strategy to continue to use symmetry as our guide. This strategy has led physicists to several compelling suggestions, which I'm sure you'll be hearing more about in future years! QCD plays an important role in all of them – either directly, as their inspiration, or as an essential tool in devising strategies for experimental exploration.

I will discuss one of these suggestions schematically, and mention three others telegraphically.

3.4.1 *Unified Field Theories*

Both QCD and the standard electroweak standard model are founded on gauge symmetries. This combination of theories gives a wonderfully economical and powerful account of an astonishing range of phenomena. Just because it is so concrete and so successful, this rendering of Nature can and should be closely scrutinized for its aesthetic flaws and possibilities. Indeed, the structure of the gauge system gives powerful suggestions for its further fruitful development. Its product structure $SU(3) \times SU(2) \times U(1)$, the reducibility of the fermion representation (that is, the fact that the symmetry does not make connections linking all the fermions), and the peculiar values of the quantum number hypercharge assigned to the known particles all suggest the desirability of a larger symmetry.

The devil is in the details, and it is not at all automatic that the superficially complex and messy observed pattern of matter will fit neatly into a simple mathematical structure. But, to a remarkable extent, it does.



	R	W	B	G	P
u	+	-	-	+	-
u	-	+	-	+	-
u	-	-	+	+	-
d	+	-	-	-	+
d	-	+	-	-	+
d	-	-	+	-	+
u^c	-	+	+	-	-
u^c	+	-	+	-	-
u^c	+	+	-	-	-
d^c	-	+	+	+	+
d^c	+	-	+	+	+
d^c	+	+	-	+	+
ν	+	+	+	+	-
e	+	+	+	-	+
e^c	-	-	-	+	+
N	-	-	-	-	-

Hypercharge Y = -1/6 (R+W+B) + 1/4 (G+P)

Figure 8: A schematic representation of the symmetry structure of the standard model. There are three independent symmetry transformations, under which the known fermions fall into five independent units (or fifteen, after threefold family repetition). The color gauge group *SU*(3) of QCD acts horizontally, the weak interaction gauge group *SU*(2) acts vertically, and the hypercharge *U*(1) acts with the relative strengths indicated by the subscripts. Right-handed neutrinos do not participate in any of these symmetries.

Figure 9: The hypothetical enlarged symmetry *SO*(10) [15] accommodates all the symmetries of the standard model, and more, into a unified mathematical structure. The fermions, including a right-handed neutrino that plays an important role in understanding observed neutrino phenomena, now form an irreducible unit (neglecting family repetition). The allowed color charges, both strong and weak, form a perfect match to what is observed. The phenomenologically required hypercharges, which appear so peculiar in the standard model, are now theoretically determined by the color and weak charges, according to the formula displayed.

Most of what we know about the strong, electromagnetic, and weak interactions is summarized (rather schematically!) in Figure 8. QCD connects particles horizontally in groups of 3 (*SU*(3)), the weak interaction connects particles vertically in groups of 2 (*SU*(2)) in the horizontal direction and hypercharge (*U*(1)) senses the little subscript numbers. Neither the different interactions, nor the different particles, are unified. There are three different interaction symmetries, and five disconnected sets of particles (actually fifteen sets, taking into account the threefold repetition of families).

We can do much better by having more symmetry, implemented by additional gluons that also change strong into weak colors. Then everything clicks into place quite beautifully, as displayed in Figure 9.

There seems to be a problem, however. The different interactions, as observed, do not have the same overall strength, as would be required by the ex-

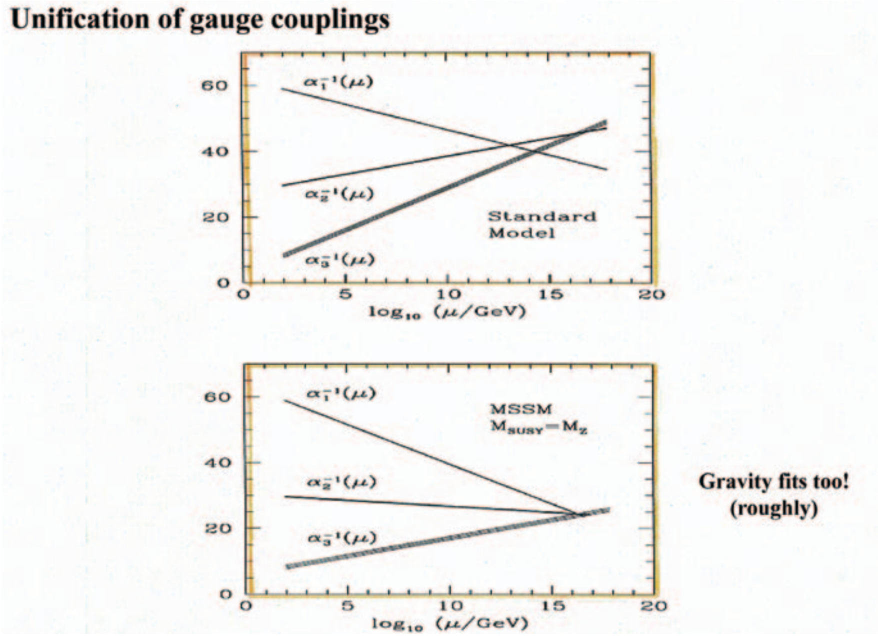


Figure 10: We can test the hypothesis that the disparate coupling strengths of the different gauge interactions derive a common value at short distances, by doing calculations to take into account the effect of virtual particle clouds [16]. These are the same sort of calculations that go into Figure 3, but extrapolated to much higher energies, or equivalently shorter distances. Top panel: using known virtual particles. Bottom panel: including also the virtual particles required by low-energy supersymmetry [17].

tended symmetry. Fortunately, asymptotic freedom informs us that the observed interaction strengths at a large distance can be different from the basic strengths of the seed couplings viewed at short distance. To see if the basic theory might have the full symmetry, we have to look inside the clouds of virtual particles, and to track the evolution of the couplings. We can do this, using the same sort of calculations that underlie Figure 3, extended to include the electroweak interactions, and extrapolated to much shorter distances (or equivalently, larger energy scales). It is convenient to display inverse couplings and work on a logarithmic scale, for then the evolution is (approximately) linear. When we do the calculation using only the virtual particles for which we have convincing evidence, we find that the couplings do approach each other in a promising way, though ultimately they don't quite meet. This is shown in the top panel of Figure 10.

Interpreting things optimistically, we might surmise from this near-success that the general idea of unification is on the right track, as is our continued reliance on quantum field theory to calculate the evolution of couplings. After all, it is hardly shocking that extrapolation of the equations for evolution of the couplings beyond their observational foundation by many orders of magnitude is missing some quantitatively significant ingredient. In a moment I'll mention an attractive hypothesis for what's missing.

A very general consequence of this line of thought is that an enormously large energy scale, of order 10^{15} GeV or more, emerges naturally as the scale of unification. This is a profound and welcome result. It is profound, because the large energy scale – which is far beyond any energy we can access directly – emerges from careful consideration of experimental realities at energies more than ten orders of magnitude smaller! The underlying logic that gives us such leverage is a synergy of unification and asymptotic freedom, as follows. If evolution of couplings is to be responsible for their observed gross inequality then, since this evolution is only logarithmic in energy, it must act over a very wide range.

The emergence of a large mass scale for unification is welcome, first, because many effects we might expect to be associated with unification are observed to be highly suppressed. Symmetries that unify $SU(3) \times SU(2) \times U(1)$ will almost inevitably involve wide possibilities for transformation among quarks, leptons, and their antiparticles. These extended possibilities of transformation, mediated by the corresponding gauge bosons, undermine conservation laws including lepton and baryon number conservation. Violation of lepton number is closely associated with neutrino oscillations. Violation of baryon number is closely associated with proton instability. In recent years neutrino oscillations have been observed; they correspond to miniscule neutrino masses, indicating a very feeble violation of lepton number. Proton instability has not yet been observed, despite heroic efforts to do so. In order to keep these processes sufficiently small, so as to be consistent with observation, a high scale for unification, which suppresses the occurrence of the transformative gauge bosons as virtual particles, is most welcome. In fact, the unification scale we infer from the evolution of couplings is broadly consistent with the observed value of neutrino masses, and that encourages further vigorous pursuit of the quest to observe proton decay.

The emergence of a large mass scale for unification is welcome, secondly, because it opens up possibilities for making quantitative connections to the remaining fundamental interaction in Nature: gravity. It is notorious that gravity is absurdly feebler than the other interactions, when they are compared acting between fundamental particles at accessible energies. The gravitational force between proton and electron, at any macroscopic distance, is about $Gm_e m_p/\alpha \sim 10^{-40}$ of the electric force. On the face of it, this fact poses a severe challenge to the idea that these forces are different manifestations of a common source – and an even more severe challenge to the idea that gravity, because of its deep connection to space-time dynamics, is the primary force.

By extending our consideration of the evolution of couplings to include gravity, we can begin to meet these challenges.

- Whereas the evolution of gauge theory couplings with energy is a subtle quantum mechanical effect, the gravitational coupling evolves even classically, and much more rapidly. For gravity responds directly to energy-momentum, and so it appears stronger when viewed with high-energy probes. In moving from the small energies where we ordinarily measure to unification energy scales, the ratio GE^2/α ascends to values that are no longer absurdly small.
- If gravity is the primary force, and special relativity and quantum mechanics frame the discussion, then Planck's system of physical units, based on Newton's constant G, the speed of light c, and Planck's quantum of action h, is privileged. Dimensional analysis then suggests that the value of naturally defined quantities, measured in these units, should be of order unity. But when we measure the proton mass in Planck units, we discover

$$m_p \sim 10^{-18}\sqrt{\frac{hc}{G}} \tag{2}$$

On this hypothesis, it makes no sense to ask "Why is gravity so feeble?". Gravity, as the primary force, just is what it is. The right question is the one we confront here: "Why is the proton so light?". Given our new, profound understanding of the origin of the proton's mass, which I've sketched for you today, we can formulate a tentative answer. The proton's mass is set by the scale at which the strong coupling, evolved down from its primary value at the Planck energy, comes to be of order unity. It is then that it becomes worthwhile to cancel off the growing color fields of quarks, absorbing the cost of quantum localization energy. In this way, we find, quantitatively, that the tiny value of the proton mass in Planck units arises from the fact that the basic unit of color coupling strength, g, is of order $\frac{1}{2}$ at the Planck scale! Thus dimensional reasoning is no longer mocked. The apparent feebleness of gravity results from our partiality toward the perspective supplied by matter made from protons and neutrons.

3.4.2 *Supersymmetry*

As I mentioned a moment ago, the approach of couplings to a unified value is suggested, but not accurately realized, if we infer their evolution by including the effect of known virtual particles. There is one particular proposal to expand the world of virtual particles, which is well motivated on several independent grounds. It is known as low-energy supersymmetry [18].

As the name suggests, supersymmetry involves expanding the symmetry of the basic equations of physics. This proposed expansion of symmetry goes in a different direction from the enlargement of gauge symmetry. Supersymmetry makes transformations between particles having the same color charges and different spins, whereas expanded gauge symmetry changes the color charges while leaving spin untouched. Supersymmetry expands the space-time symmetry of special relativity.

In order to implement low-energy supersymmetry, we must postulate the existence of a whole new world of heavy particles, none of which has yet been observed directly. There is, however, a most intriguing indirect hint that this idea may be on the right track: If we include the particles needed for low-energy supersymmetry, in their virtual form, in the calculation of how couplings evolve with energy, then accurate unification is achieved! This is shown in the bottom panel of Figure 10.

By ascending a tower of speculation, involving now both extended gauge symmetry and extended space-time symmetry, we seem to break though the clouds, into clarity and breathtaking vision. Is it an illusion, or reality? This question creates a most exciting situation for the Large Hadron Collider (LHC), due to begin operating at CERN in 2007, for this great accelerator will achieve the energies necessary to access the new world of of heavy particles, if it exists. How the story will play out, only time will tell. But in any case I think it is fair to say that the pursuit of unified field theories, which in past (and many present) incarnations has been vague and not fruitful of testable consequences, has in the circle of ideas I've been describing here attained entirely new levels of concreteness and fecundity.

3.4.3 *Axions* [19]

As I have emphasized repeatedly, QCD is in a profound and literal sense constructed as the embodiment of symmetry. There is an almost perfect match between the observed properties of quarks and gluons and the most general properties allowed by color gauge symmetry, in the framework of special relativity and quantum mechanics. The exception is that the established symmetries of QCD fail to forbid one sort of behavior that is not observed to occur. The established symmetries permit a sort of interaction among gluons – the so-called θ term – that violates the invariance of the equations of QCD under a change in the direction of time. Experiments provide extremely severe limits on the strength of this interaction, much more severe than might be expected to arise accidentally.

By postulating a new symmetry, we can explain the absence of the undesired interaction. The required symmetry is called Peccei-Quinn symmetry after the physicists who first proposed it. If it is present, this symmetry has remarkable consequences. It leads us to predict the existence of new very light, very weakly interacting particles, *axions.* (I named them after a laundry detergent, since they clean up a problem with an axial current.) In principle axions might be observed in a variety of ways, though none is easy. They have interesting implications for cosmology, and they are a leading candidate to provide cosmological dark matter.

3.4.4 *In Search of Symmetry Lost* [20]

It has been almost four decades since our current, wonderfully successful theory of the electroweak interaction was formulated. Central to that theory is the concept of spontaneously broken gauge symmetry. According to this concept, the fundamental equations of physics have more symmetry than the actual physical world does. Although its specific use in electroweak theory involves exotic hypothetical substances and some sophisticated mathematics, the underlying theme of broken symmetry is quite old. It goes back at least to the dawn of modern physics, when Newton postulated that the basic laws of mechanics exhibit full symmetry in three dimensions of space despite the fact that everyday experience clearly distinguishes 'up and down' from 'sideways' directions in our local environment. Newton, of course, traced that asymmetry to the influence of Earth's gravity. In the framework of electroweak theory, modern physicists similarly postulate that the physical world is described by a solution wherein all space, throughout the currently observed Universe, is permeated by one or more (quantum) fields that spoil the full symmetry of the primary equations.

Fortunately this hypothesis, which might at first hearing sound quite extravagant, has testable implications. The symmetry-breaking fields, when suitably excited, must bring forth characteristic particles: their quanta. Using the most economical implementation of the required symmetry breaking, one predicts the existence of a remarkable new particle, the so-called Higgs particle. More ambitious speculations suggest that there should be not just a single Higgs particle, but rather a complex of related particles. Low-energy supersymmetry, for example, requires at least five "Higgs particles".

Elucidation of the Higgs complex will be another major task for the LHC. In planning this endeavor, QCD and asymptotic freedom play a vital supporting role. The strong interaction will be responsible for most of what occurs in collisions at the LHC. To discern the new effects, which will be manifest only in a small proportion of the events, we must understand the dominant backgrounds very well. Also, the production and decay of the Higgs particles themselves usually involves quarks and gluons. To anticipate their signatures, and eventually to interpret the observations, we must use our understanding of how protons – the projectiles at LHC – are assembled from quarks and gluons, and how quarks and gluons show themselves as jets.

4 THE GREATEST LESSON

Evidently asymptotic freedom, besides resolving the paradoxes that originally concerned us, provides a conceptual foundation for several major insights into Nature's fundamental workings, and a versatile instrument for further investigation.

The greatest lesson, however, is a moral and philosophical one. It is truly awesome to discover, by example, that we humans can come to comprehend Nature's deepest principles, even when they are hidden in remote and alien realms. Our minds were not created for this task, nor were appropriate tools ready at hand. Understanding was achieved through a vast international effort involving thousands of people working hard for decades, competing in the small but cooperating in the large, abiding by rules of openness and honesty. Using these methods – which do not come to us effortlessly, but require nurture and vigilance – we can accomplish wonders.

5 POSTCRIPT: REFLECTIONS

That was the conclusion of the lecture as I gave it. I'd like to add, in this written version, a few personal reflections.

5.1 *Thanks*

Before concluding I'd like to distribute thanks.

First I'd like to thank my parents, who cared for my human needs and encouraged my curiosity from the beginning. They were children of immigrants from Poland and Italy, and grew up in difficult circumstances during the Great Depression, but managed to emerge as generous souls with an inspiring admiration for science and learning. I'd like to thank the people of New York, for supporting a public school system that served me extremely well. I also got a superb undergraduate education, at the University of Chicago. In this connection I'd especially like to mention the inspiring influence of Peter Freund, whose tremendous enthusiasm and clarity in teaching a course on group theory in physics was a major influence in nudging me from pure mathematics toward physics.

Next I'd like to thank the people around Princeton who contributed in crucial ways to the circumstances that made my development and major work in the 1970s possible. On the personal side, this includes especially my wife Betsy Devine. I don't think it's any coincidence that the beginning of my scientific maturity, and a special surge of energy, happened at the same time as I was falling in love with her. Also Robert Shrock and Bill Caswell, my fellow graduate students, from whom I learned a lot, and who made our extremely intense life-style seem natural and even fun. On the scientific side, I must of course thank David Gross above all. He swept me up in his drive to know and to calculate, and through both his generous guidance and his personal example started and inspired my whole career in physics. The environment for theoretical physics in Princeton in the 1970s was superb. There was an atmosphere of passion for understanding, intellectual toughness, and inner confidence whose creation was a great achievement. Murph Goldberger, Sam Treiman, and Curt Callan especially deserve enormous credit for this. Also Sidney Coleman, who was visiting Princeton at the time, was very actively interested in our work. Such interest from a physicist I regarded as uniquely brilliant was inspiring in itself; Sidney also asked many challenging specific questions that helped us come to grips with our results as they developed. Ken Wilson had visited and lectured a little earlier, and his renormalization group ideas were reverberating in our heads.

Fundamental understanding of the strong interaction was the outcome of decades of research involving thousands of talented people. I'd like to thank my fellow physicists more generally. My theoretical efforts have been inspired by, and of course informed by, the ingenious persistence of my experimental colleagues. Thanks, and congratulations, to all. Beyond that generic thanks I'd like to mention specifically a trio of physicists whose work was particularly important in leading to ours, and who have not (yet?) received a Nobel Prize for it. These are Yoichiro Nambu, Stephen Adler, and James Bjorken. Those heroes advanced the cause of trying to understand hadronic physics by taking the concepts of quantum field theory seriously, and embodying them in specific mechanistic models, when doing so was difficult and unfashionable. I'd like to thank Murray Gell-Mann and Gerard 't Hooft for not quite inventing everything, and so leaving us something to do. And finally I'd like to thank Mother Nature for her extraordinarily good taste, which gave us such a beautiful and powerful theory to discover.

This work is supported in part by funds provided by the U.S. Department of Energy (D.O.E.) under cooperative research agreement DE-FC02-94ER40818.

5.2 *A Note to Historians*

I have not, here, given an extensive account of my personal experiences in discovery. In general, I don't believe that such accounts, composed well after the fact, are reliable as history. I urge historians of science instead to focus on the contemporary documents; and especially the original papers, which by definition accurately reflect the understanding that the authors had at the time, as they could best articulate it. From this literature, it is I think not difficult to identify where the watershed changes in attitude I mentioned earlier occurred, and where the outstanding paradoxes of strong interaction physics and quantum field theory were resolved into modern paradigms for our understanding of Nature.

REFERENCES

[1] In view of the nature and scope of this write-up, its footnoting will be light. Our major original papers [2, 3, 4] are carefully referenced.
[2] D. Gross and F. Wilczek, *Phys. Rev. Lett.* **30**, 1343 (1973).
[3] D. Gross and F. Wilczek, *Phys. Rev.* **D8**, 3633 (1973).
[4] D. Gross and F. Wilczek, *Phys. Rev.* **D9**, 980 (1974).
[5] J. Friedman, H. Kendall, and R. Taylor received the Nobel Prize for this work in 1990.
[6] L. Landau, in *Niels Bohr and the Development of Physics,* ed. W. Pauli (McGraw-Hill, New York 1955).
[7] C.-N. Yang and R. Mills, *Phys. Rev.* **96**, 191 (1954).
[8] An especially clear and insightful early paper, in which a dynamical role for color was proposed, is Y. Nambu, in *Preludes in Theoretical Physics* ed. A. De-Shalit, H. Feshbach, L. van Hove (North-Holland, Amsterdam 1966).
[9] Figure courtesy L3 collaboration, CERN.
[10] Figure courtesy S. Bethke, hep-ex/0211012.
[11] Figure courtesy Center for Computational Physics, University of Tsukuba.
[12] Figure courtesy D. Leinweber, http://www.physics.adelaide.edu.au/theory/staff/leinweber/VisualQCD/Nobel/.

[13] Figure courtesy G. Kilcup, http://www.physics.ohio-state.edu/~kilcup.
[14] Figure courtesy STAR collaboration, Brookhaven National Laboratory.
[15] Unification based on *SO*(10) symmetry was first outlined in H. Georgi, in *Particles and Fields - 1974,* ed. C. Carlson (AIP, New York, 1975).
[16] H. Georgi, H. Quinn, and S. Weinberg, *Phys. Rev. Lett.* **33**, 451 (1974).
[17] S. Dimopoulos, S. Raby, and F. Wilczek, *Phys. Rev.* **D24**, 1681 (1981).
[18] A standard review is H. P. Nilles, *Phys. Rep.* **110**, 1 (1984).
[19] A standard review is J. Kim, *Phys. Rep.* **150**, 1 (1987). I also recommend F. Wilczek, hep-ph/0408167.
[20] I treat this topic more amply in F. Wilczek, *Nature* **433**, 239 (2005).

THE DISCOVERY OF ASYMPTOTIC FREEDOM AND THE EMERGENCE OF QCD

Nobel Lecture, December 8, 2004

by

David J. Gross

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

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 PHYSCIS IN THE 1960'S

The early 1960's, 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 non-perturbative tools. By the 1950's 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. Tomanaga and F. Dyson, which had eliminated the ubiquitous infinities that occurred 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 1940's.

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 [1], *"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 1950's had failed. The path integral, developed by Feynman in the late 1940's, 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 non-perturbative 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 1950's 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 non-perturbative 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 non-renormalizable 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 1950's, 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 1950's, 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."* [2].

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"*. [3]. 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"*. [4].

Under the influence of Landau and Pomeranchuk, a generation of physicists was forbidden to work on field theory. Why the discovery of the zero charge problem did 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. Terentev carried out a calculation of the charge renormalization of charged vector mesons in 1964 [5]. 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 non-renormalizability 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 [6]. 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 conflicts 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 NAL 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".* [7].

SYMMETRIES

If dynamics was forbidden, one could at least explore the symmetries of the strong interactions. The biggest advance of the early 1960's was the discovery of an approximate symmetry of hadrons, SU(3), by Gell-Mann and Y. Neeman [8], 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 ices of veal, which are then discarded".* [9]. This paper made quite an impression, 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, 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) [10]. 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 [11], Y. Nambu [12], M.Y. Han and Nambu [13]. Nambu's motivation for color was two-fold; 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 with 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 was ten times the threshold for their production. The non-relativistic 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 multi-body 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".* [14]. 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, but 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 [15]. 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 [16]. 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 ones' 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 1967 paper, on the application of U(6)XU(6), particularly influenced me[17]. 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 [18].

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 [19]. 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 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 $\sigma_L = 0$ [20]. 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 point-like objects of spin one-half. In the spring of 1969, which I spent at CERN, C. Llewelynn-Smith and I analyzed the sum rules that followed for deep-inelastic neutrino-nucleon scattering using similar methods [21]. 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 point-like constituents of the nucleon. But how could that be? Surely strong interactions must exist between the quarks that would smear out their point-like behavior.

After the experiments at SLAC, Feynman came up with his parton picture of deep inelastic scattering [22], a very picturesque and intuitive way of describing deep-inelastic scattering in terms of assumed point-like 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 [23]. 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 [24]. However, I soon realized that in a field theoretic context only a free, non-interacting 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. Symanzik) had rediscovered the renormalization group equations, which they presented as a consequence of a scale invariance anomaly [25]. 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. But 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 turn of events in the history of physics. In 1968 G. Veneziano came up with a remarkably simple formula that summarized many features of hadronic scattering [26], with Regge asymptotic behavior in one channel and narrow resonance saturation in the other. This formula was soon generalized to multi-particle 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 lead to incredibly soft behavior at large momentum transfer, quite the opposite of the hard scaling observed. Also, it required for consistency many 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 may 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 cannot 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 non-canonical scaling indicative of a non-trivial fixed point of the renormalization group would appear at higher energies.

Much happened during the next two years. Gerhard 't Hooft's spectacular work on the renormalizability of Yang-Mills theory reintroduced non-Abelian gauge theories to the community [27]. 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 [28] provided a tool that could be applied to the analysis of deep inelastic scattering. Wilson's development of the operator product expansion 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 [29]. The operator product analysis was extended to the light cone, the relevant region for deep-inelastic scattering [30]. Most important from my point of view was the revival of the renormalization group by Wilson [31]. The renormalization group stems from the fundamental work of Gell-Mann and Low, Stueckelberg and Petermann, and Bogoliubov and Shirkov [32]. This work was neglected for many years, partly because it seemed to provide only information about physics for large space-like 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 point-like 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 light-like 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 [33] 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 naïve 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 [34]. 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 [35]. 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 [36].

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 working 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 electro-weak 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 me 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 analogue, 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, where 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, non-asymptotically 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 [37,38].

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 [39]. In a relativistic theory one can calculate the dielectric constant, ε, in terms of the magnetic permeability, μ, since $\varepsilon\mu=1$ (in units where c=velocity of light=1). In classical physics all media are diamagnetic. This is because, classsically, 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 $\varepsilon>1$. 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 one. 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 anti-screening of the Yang-Mills vacuum as paramagnetism. QCD is asymptotically free because the anti-screening 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 $\delta\mu = (-1/3+2^2)q^2 = 11/3\ q^2$; whereas a spin one-half quark contributes, $\delta\mu = -(-1/3+(2/2)^2)\ q^2 = -2/3\ q^2$ (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 anti-screening 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:

$$\beta(\alpha) \equiv \mu \frac{d\alpha(\mu)}{d\mu} = \frac{\alpha^2}{\pi} b_1 + \frac{\alpha^3}{\pi^2} b_2 + \ldots \quad \text{, where} \quad \alpha = \frac{g^2}{4\pi} \tag{1}$$

Our result was that [37,38]

$$b_1 = -\left[\frac{11}{6} C_A - \frac{2}{3} \Sigma_R n_R T_R\right] \tag{2}$$

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=N, T_F=1/2, and thus

$$b_1 = -11/2 + n_F/3. \tag{3}$$

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 important of all the suggestion that the strong interactions must be based on a color gauge theory [37].

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 three) that went into the evaluation of the $\pi \rightarrow 2\gamma$ decay rate from the axial anomaly (this had been recently emphasized by W. Bardeen, H. Fritzsch and Gell-Mann [40]), and the factor of three that color provided in the total annihilation cross section. Thus the gluons could couple to color and all would be well. Thus we proposed [37]: *"One particularly appealing model is based on three triplets of fermions, with Gell-Mann's SU(3)xSU(3)as a global symmetry and a 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)x 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 [34]. 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 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"* [37]. Politizer'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 [38].

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 [41]. 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 [42]. 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 naïve 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."* [41].

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 found that these effects do not violate conservation of parity and strangeness, in agreement with observation, as long as there were no colored scalars [43]. 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 non-confining, 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, Fritzsch, Gell-Mann and H. Leutwyler submitted a paper in which they discussed the *"advantages of 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"* [44]. They discussed various models and pointed out the advantages of each. The first point was already discussed at the NAL high-energy physics conference in August 1972. There Gell-Mann and Fritzsch had discussed their program of *"abstracting results from the quark-gluon model"* [45]. They discussed various models and asked, *"shall treat the vector gluon, for convenience, as a color singlet"*. In October 1973, Fritzsch, Gell-Mann and Leutwyler also noted that in the non-relativistic 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 [46]. 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 QCD first appeared in a review by W. Marciano and H. Pagels [47], 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)x 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, R. Gatto, Georgi, Glashow, Kogut, Ben Lee, Maiani, Migdal, Polyakov, Politzer, Susskind, Weinberg, 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 [48] 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 now 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 [49]. 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 [50]. Two dimensional gauge theories trivially confine color. This was realized quite early and discussed for Abelian gauge theory, the Schwinger model, by A. Casher, Kogut and Susskind, as a model for confinement in the fall of 1973 [51]. However, QCD_2 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 (2), $b_2 = -\left[\frac{17}{12}C_A^{\ 2} - \frac{1}{2}C_F T_F n - \frac{5}{6}C_A T_F n\right]$ [52]. Appelquist and Georgi, and Zee calculated the corrections to the scaling of the e^+-e^- annihilation cross-section[53]; Gaillard, and Lee, and independently Altarelli and Maiani, calculated the enhancement of the $\Delta I=1/2$ non-leptonic decay matrix elements [54]. The analysis of scaling violations for deep-inelastic scattering continued [55], 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 $R = \sigma_{e^+e^- \to hadrons} / \sigma_{e^+e^- \to \mu^+\mu^-}$ 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 $R \to 3[(1/3)^2 + (1/3)^2 + (2/3)^2] = 2$. 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 GIM mechanism to work, but as C. Bouchiat, J. Illiopoulos and L. Maini, and independently R. Jackiw and I [56] showed, if the charmed quark were absent the electro-weak theory would be anomalous and non-renormalizable. Gaillard, Lee and Rosner had written an important and insightful paper on the phenomenology of charm [57]. 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 [58]), 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 1975–78 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 is believing. We now believe in the physical reality of quarks and gluons, we now believe in 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 point like 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 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 electro-weak 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 operates 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, where nucleons melt and quarks and gluons are liberated to form a quark-gluon plasma.

d. Unification.

One the most important implication 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 strength of the strong interactions and the electro-weak 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. H. Georgi, H. Quinn and S. Weinberg showed that the couplings run in such a way as to merge somewhere around 10^{14} to 10^{16} Ge*v* [59]. This is our most direct clue as to where the next threshold of fundamental physics lies, and hints that at this immense energy all the forces of nature, including gravity, are unified.

ACKNOWLEDGEMENTS

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

1. Feynman, R., in *The Quantum Theory of Fields* – The 12th Solvay Conference, 1961 (Interscience, New York).
2. Landau, L.D. and I.Pomeranchuk, *Dokl. Akad. Nauk SSSR* **102**, 489 (1955).
3. Landau, L.D., in *Niels Bohr and the Development of Physics,* (ed. W.Pauli), p. 52. (Pergamon Press, London, 1955).
4. Landau, L.D., Fundamental Problems, in *Pauli Memorial Volume,* pg. 245, (Interscience, New York, 1960).
5. VanTerent'ev, M.V. and Vanyashin, V.S., *Soviet Physics JETP* **21**, 380 (1965).
6. Chew, G. in *S-Matrix Theory,* (W.A. Benjamin Inc, 1963).
7. Gell-Mann, M., Current Algebra: Quarks and What Else?, in *Proceedings of the XVI Int'l Conference on High Energy Physics,* Vol. **4**, 135 (1972).
8. Gell-Mann, M. and Neeman, Y. *The Eightfold Way* (W.A. Benjamin Inc.) (1964).
9. Gell-Mann, M., The Symmetry Group of Vector and Axial Vector Currents, *Physics,* **1**, 63 (1964).
10. Gell-Mann, M., A Schematic Model of Baryons and Mesons, *Phys. Lett.* **8**, 214 (1964); Zweig, G., *CERN Report No. TH401,* 4R12 (1964) (unpublished).
11. Greenberg, O.W., Spin and Unitary-Spin Independence in a Paraquark Model of Baryons and Mesons, *Phys. Rev. Lett.* **13**, 598 (1964).
12. Nambu,Y., Dynamical Symmetries & Fundamental Fields, *Proceedings of 2nd Coral Gables Conf. on Symmetry Principles at High Energy,* 133(1965); Systematics of Hadrons in Subnuclear Physics, *Preludes in Theoretical Physics,* (1965).
13. Han, M.Y. and Nambu,Y., Three-Triplet Model with Double SU(3) Symmetry, *Phys. Rev.* **139 B**, 1006 (1965).
14. Low, F., in *Proceedings of the International Conference on High Energy Physics 1966,* pg. 249 (1966).
15. Adler, S., Sum Rules for the Axial Vector Coupling Constant Renormalization in Beta Decay, *Phys. Rev.* **140 B**, 736 (1965); Weisberger,W.I. Unsubtracted Dispersion Relations and the Renormalization of the Weak Axial-Vector Coupling Constants, *Phys. Rev.* **143 B**, 1302 (1966).
16. Dashen, R. and Gell-Mann, M., Representations of Local Current Algebra at Infinite Momentum, *Phys. Rev. Lett.* **17**, 340 (1966).
17. Bjorken, J.D. Applications of the Chiral U(6)XU(6)Algebra of Current Densities, *Phys. Rev.* **148 B**, 1467 (1966); Current Algebra at Small Distances, *Varenna School Lectures,* Course XLI, Varenna, Italy (1967).
18. Callan, C.G. and Gross, D. J., Crucial Test of a Theory of Currents, *Phys. Rev. Lett* **21**, 311 (1968).
19. Bloom, E.D. *et al.*, High-Energy Inelastic e-p Scattering at 6° and 10°, P*hys. Rev. Lett.* **23**, 930 (1969).
20. Callan, C. G. and Gross, D. J., High-Energy Electroproduction and the Constitution of the Electric Current, *Phys. Rev. Lett.* **22**, 156 (1968).
21. Gross, D.J. and Llewelyn-Smith,C., High Energy Neutrino-Nucleon Scattering, Current Algebra and Partons, *Nucl. Phys.* **B14**, 337 (1969).
22. Feynman, R.P., High-Energy Collisions of Hadrons, *Phys.Rev.Lett.* **23**, 1415 (1969).
23. Drell, S.D. and Yan, T.M., Partons and Their Applications at High Energies, *Ann. Phys.* (NY) **66**, 578 (1971).
24. Gross, D.J. and Wess, J., Scale Invariance, Conformal Invariance and the High Energy Behavior of Scattering Amplitudes, *Phys. Rev.* **D2**, 753 (1970).
25. Callan, C.G., Broken Scale Invariance in Scalar Field Theory, *Phys. Rev.* **D2**, 1541 (1970); Symansik, K., Small Distance Behavior in Field Theory and Power Counting, *Comm. Math. Phys.* **18**, 227 (1970).
26. Veneziano, G., Construction of a Crossing-Symmetric Regge-Behaved Amplitude for Linearly Rising Trajectories, *Nuovo Cimento,* **57 A**, 190 (1968).

27. ‘t Hooft, G., Renormalizable Lagrangians for Massive Yang-Mills Fields, *Nucl. Phys.* **35**, 167 (1967).
28. Wilson, K., Renormalization Group and Strong Interactions, *Phys. Rev.* **D3**, 1818 (1971). 29.
29. Callan, C.G., Broken Scale Invariance and Asymptotic Behavior, *Phys. Rev.* **D5**, 3202 (1972); Symansik, K., Small-Distance-Behavior Analysis and Wilson Expansions, *Comm. Math. Phys.* **23**, 49 (1971); Christ, N., Hasslacher, B. and Mueller, A., Light-Cone Behavior of Perturbation Theory, *Phys. Rev.* **D6**, 3543 (1972); Callan, C. G .and Gross, D. J., Bjorken Scaling in Quantum Field Theory, *Phys. Rev.* **D8**, 4383 (1973).
30. Jackiw, R. Van Royen, R. and West, G., Measuring Light-Cone Singularities, *Phys. Rev.* **D2**, 2473 (1970); Frishman, Y., Operator Products at Almost Light Like Distances, *Ann. Phys.* **66**, 373 (1971); Leutwyler, H. and Stern, J., Singularities of Current Commutators on the Light Cone, *Nucl. Phys.* **B20**, 77 (1970); Gross, D. J., Light Cone Structure of Current Commutators in the Gluon-Quark Model, *Phys. Rev.* **D4**, 1059 (1971); Christ, N., Hasacher, B. and Mueller, A., Light-Cone Behavior of Perturbation Theory, *Phys. Rev.* **D6**, 3543 (1972).
31. Wilson, K. and Kogut, J., The Renormalization Group and the ε expansion. *Phys. Rep.* **31**, 792 (1973).
32. Gell-Mann, M. and Low, F., Quantum Electrodynamics at Small Distances, *Phys. Rev.* **95**, 1300 (1954). Stueckelberg, E. and Petermann, A., La Normalisation des Constantes dans la Theorie des Quanta, *Helv. Phys. Acta* **26**, 499 (1953). N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields,* Interscience, New York, (1959).
33. Parisi, G., Deep Inelastic Scattering in a Field Theory with Computable Large-Momenta Behavior, *Lett. al Nuovo Cimento* **7**, 84 (1973).
34. Callan, C.G. and Gross, D. J., Bjorken Scaling in Quantum Field Theory, *Phys. Rev.* **D8**, 4383 (1973).
35. Coleman, S. and Gross, D. J., Price of Asymptotic Freedom, *Phys. Rev. Lett.* **31**, 851 (1973).
36. Zee, A., Study of the Renormalization Group for Small Coupling Constants, *Phys. Rev.* **D7**, 3630 (1973).
37. Gross, D. J. and Wilczek,F., Ultraviolet Behavior of Non-Abelian Gauge Theories, *Phys. Rev. Lett.* **30**, 1343 (1973).
38. Politzer, H.D., Reliable Perturbative Results for Strong Interactions?, *Phys. Rev. Lett.* **30**, 1346 (1973).
39. Nielsen, N.K., Asymptotic Freedom as a Spin Effect, *Am.J. Phys.* **49**, 1171(1981).
40. Bardeen, W.A., Fritzsch, H. and Gell-Mann, M., Light-Cone Current Algebra, π^0 Decay, and e^+-e^- Annihilation, in *Scale and Conformal Symmetry in Hadron Physics,* ed. R. Gatto (Wiley & Sons, 1973).
41. Gross, D. J. and Wilczek, F., Asymptotically Free Gauge Theories. I, *Phys. Rev.* **D8**, 3633 (1973).
42. Gross, D. J. and Wilczek, F., Asymptotically Free Gauge Theories. II, *Phys. Rev.* **D9**, 980 (1974).
43. Weinberg, S., Non-Abelian Gauge Theories of the Strong Interactions, *Phys. Rev. Lett.* **31**, 494 (1973).
44. Fritzsch, H., Gell-Mann, M. and Leutwyler, H., Advantages of the Color Octet Gluon Picture, *Phys. Lett.* **47 B**, 368 (1973).
45. Fritzsch,H. and Gell-Mann, M., Current Algebra: Quarks and What Else?, *Proceedings of the XVI Int. Conf. on High Energy Physics,*Vol.2,164 (1972).
46. Lipkin, H., Triality, Exotics and the Dynamical Basis of the Quark Model, *Phys. Lett.* **45 B**, 267 (1973).
47. Marciano,W. and Pagels, H., Quantum Chromodynamics, *Phys. Rep.* **C36**, 137 (1978).
48. Wilson, K., Confinement of Quarks, *Phys. Rev.* **D10**, 2445 (1974).
49. Gross, D. J. and Neveu, A., Dynamical Symmetry Breaking in an Asymptotically Free Theory, *Phys. Rev.* **D10**, 3235 (1974).
50. t’ Hooft, G., A Planar Diagram Theory for Strong Interactions, *Nucl. Phys.* **B72**, 461 (1974).

51. A. Casher, J. Kogut and L. Susskind, Vacuum Polarization and the Quark-Parton puzzle, *Phys. Rev. Lett.* **31**, 792 (1973); Vacuum Polarization and the Absence of Free Quarks, *Phys. Rev.* **D10**, 732 (1974).
52. Caswell, W., Asymptotic Behavior of Non-Abelian Gauge Theories to Two-Loop Order, *Phys. Rev. Lett.* **33**, 244 (1974); Jones, D., Two-Loop Diagrams in Yang-Mills Theory, *Nucl. Phys.* **B75**, 531, (1974).
53. Appelquist, T. and Georgi, H., e^+-e^- Annihilation in Gauge Theories of Strong Interactions, *Phys. Rev.* **D8**, 4000 (1973); Zee, A., Electron-Positron Annihilation in Stagnant Field Theories, *Phys. Rev.* **D8**, 4038, (1973).
54. Gaillard, M.K. and Lee, B.W. ΔI=1/2 Rule for Non-Leptonic Decays in Asymptotically Free Field Theories, *Phys. Rev. Lett.* **33**, 108 (1974); Altarelli, G. and Maiani,L., Octet Enhancement of Non-Leptonic Weak Interactions in Asymptotically Free Gauge Theories *Phys. Lett.* **52 B**, 351 (1974).
55. Gross, D. J., How to Test Scaling in Asymptotically Free Theories, *Phys. Rev. Lett.* **32**, 1071 (1974).
56. Bouchiat, C., Iliopoulpos, J. and Meyer, Ph., An Anomaly Free Version of Weinberg's Model, *Phys. Lett.* **38 B**, 519 (1972); Gross, D.J. and Jackiw, R., The Effect of Anomalies on Quasi-Renormalizable Theories, *Phys. Rev.* **D6**, 477 (1972).
57. Gaillard, M.K., Lee, B.W. and Rosner, J.L., Search for Charm, *Rev. Mod. Phys.* **47**, 277 (1975).
58. Appelquist, T. and Politzer, H. D., Heavy Quarks and e^+-e^- Annihilation, *Phys. Rev. Lett.* **34**, 43 (1975).
59. Georgi, H., Quinn, H.R. and Weinberg, S., Hierarchy of Interactions in Unified Gauge Theories, *Phys. Rev. Lett.* **33**, 451 (1974).

Asymptotic Freedom:

From Paradox to Paradigm

Paradox 1:

Quarks are Born Free,
But Everywhere They
are in Chains

The quark model “works” …

… but its rules are very strange.

Quarks behave independently when they're close, but they can't be pulled apart. An unprecedented hypothesis: CONFINEMENT!

Hard-hit quarks accelerate rapidly, without radiating away energy. The strongest force of nature "turns off": FREEDOM!

Paradox 2:

Quantum Mechanics and Special Relativity Both Work

Special relativity puts space and time on the same footing, but quantum mechanics treats them very differently. This leads to a creative tension ...

Dirac: from uncertainty to antiparticles

Feynman-Schwinger-Tomonoga: the reality of virtual particles (QED)

'tHooft-Veltman: the vast scope of virtual particles (electroweak gauge theory)

Landau's Paradox

Screening by virtual particles wipes out interactions

The demise of quantum field theory was widely proclaimed - and welcomed!

Paradox Lost:

Asymptotic Freedom

Some very special quantum field theories have **anti**-screening (asymptotic freedom).

One of these theories is uniquely well suited to accommodating quarks. It **predicts** gluons. This is quantum chromodynamics (QCD).

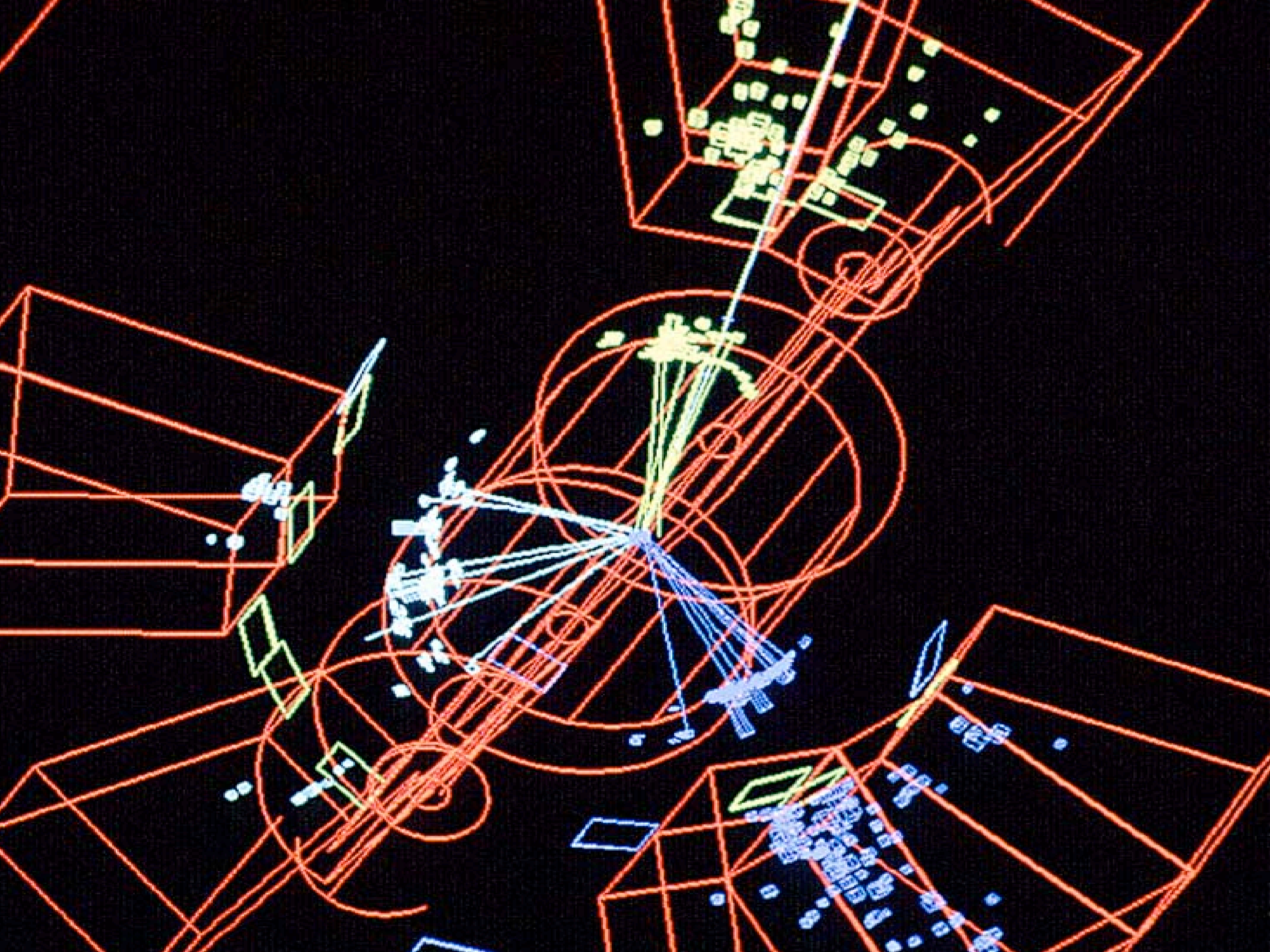
Antiscreening explains how the same basic interaction can appear either powerful or feeble, depending on circumstances.

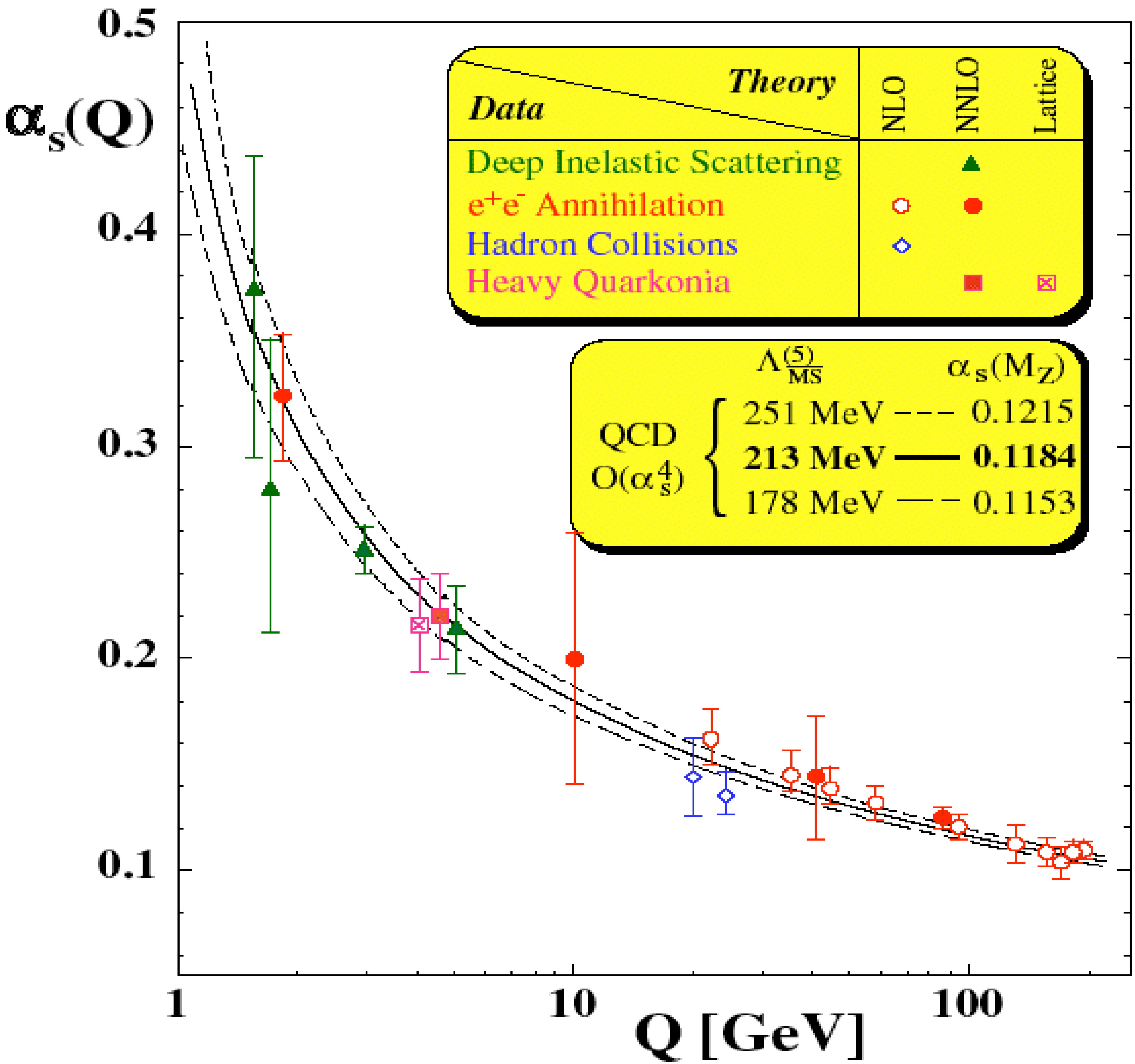
The interaction is feeble at small separations, powerful at large separations. (Confinement!)

The interaction does not interfere with violent deflections. (Freedom!) Nor does it induce them.

Paradigm 1:

The Hard Reality of Quarks and Gluons





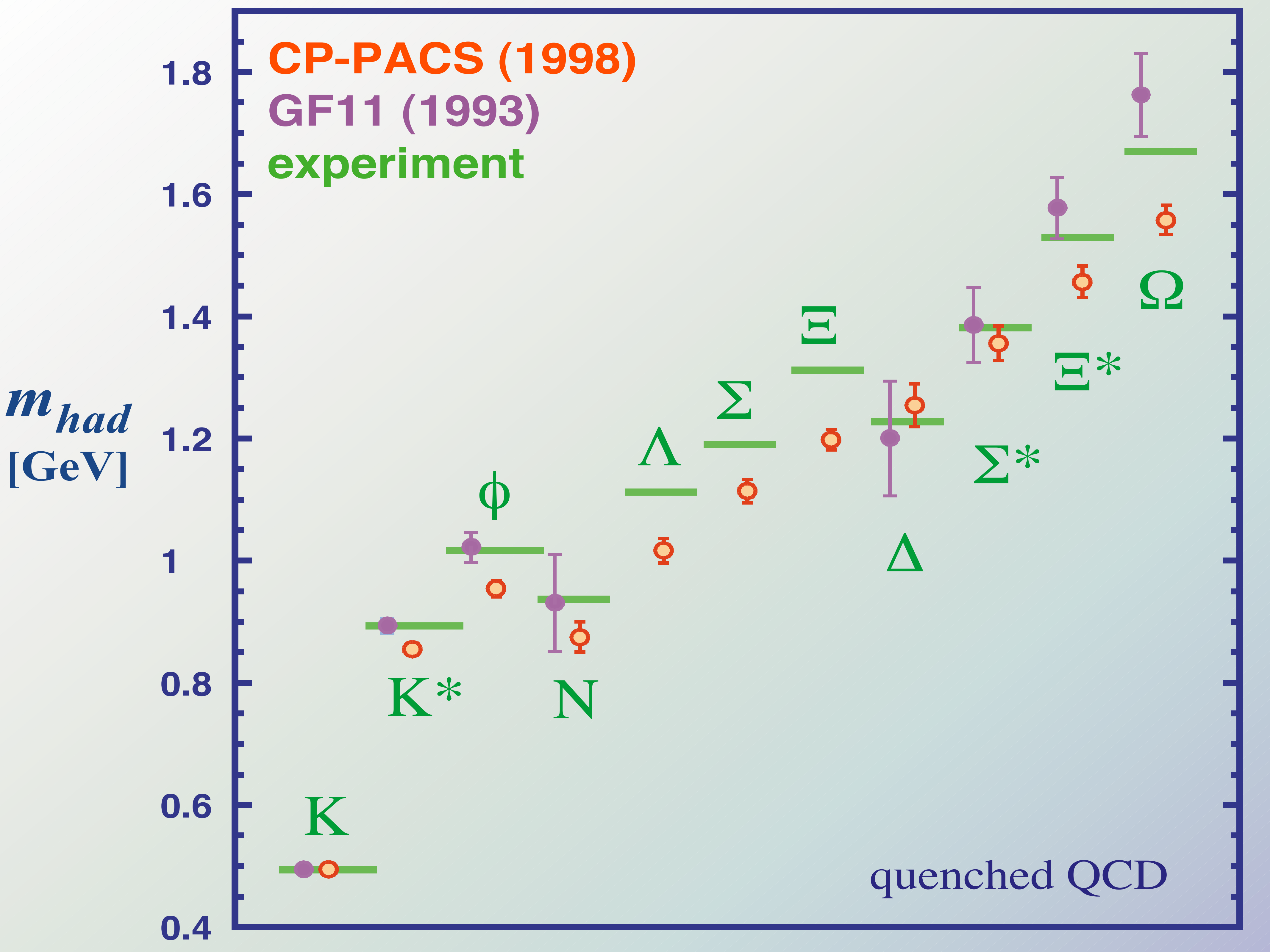
0.5
0.4
0.3
0.2
0.1
$\alpha_s(Q)$
Theory
Data
NLO
NNLO
Lattice
Deep Inelastic Scattering
e^+e^- Annihilation
Hadron Collisions
Heavy Quarkonia
$\Lambda^{(5)}_{\overline{MS}}$
$\alpha_s(M_Z)$
QCD
$O(\alpha_s^4)$
251 MeV 0.1215
213 MeV 0.1184
178 MeV 0.1153
1
10
100
Q [GeV]

Paradigm 2:

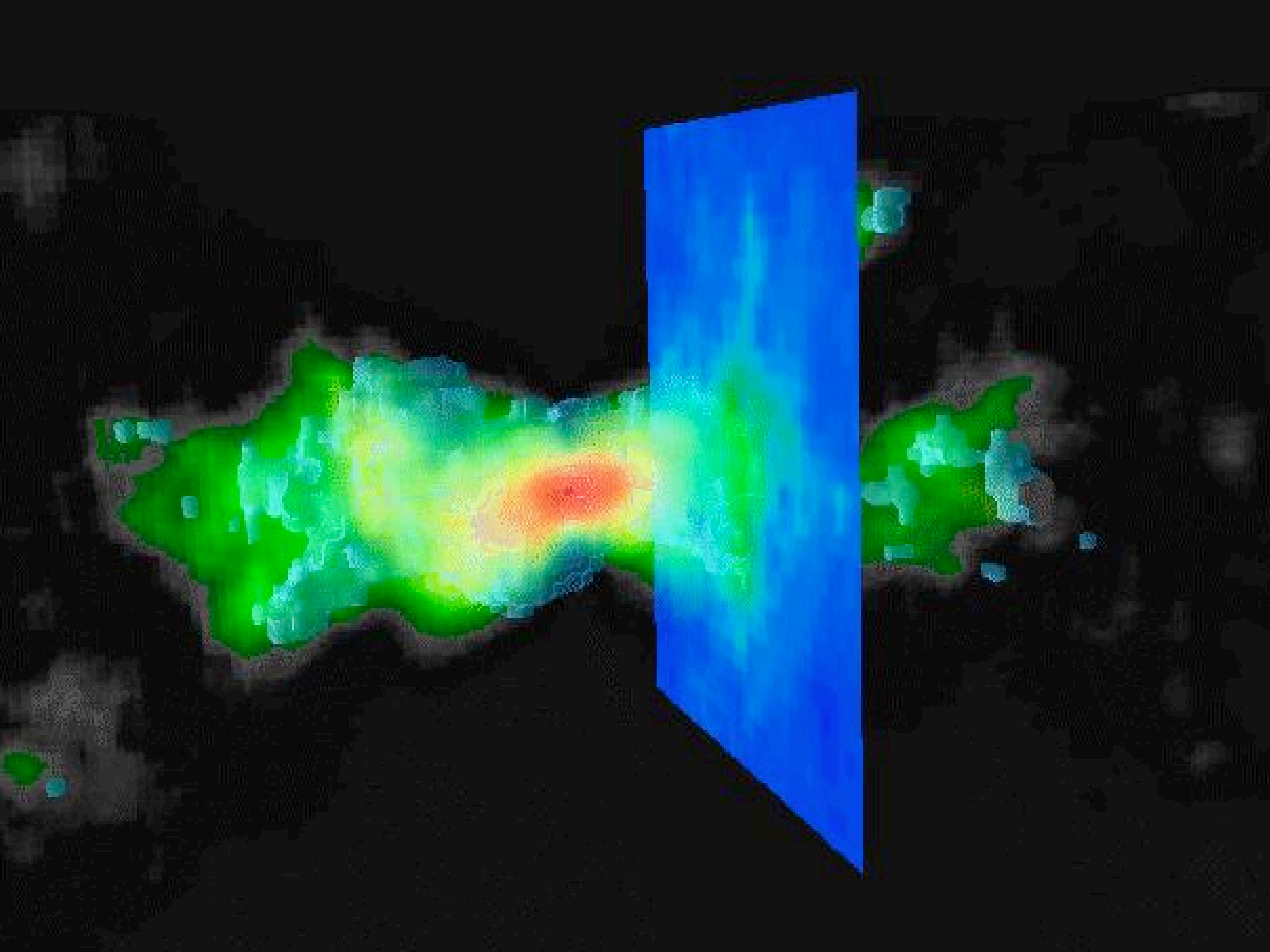
Mass Comes From Energy

Einstein's Second Law:

$$m=E/c^2$$

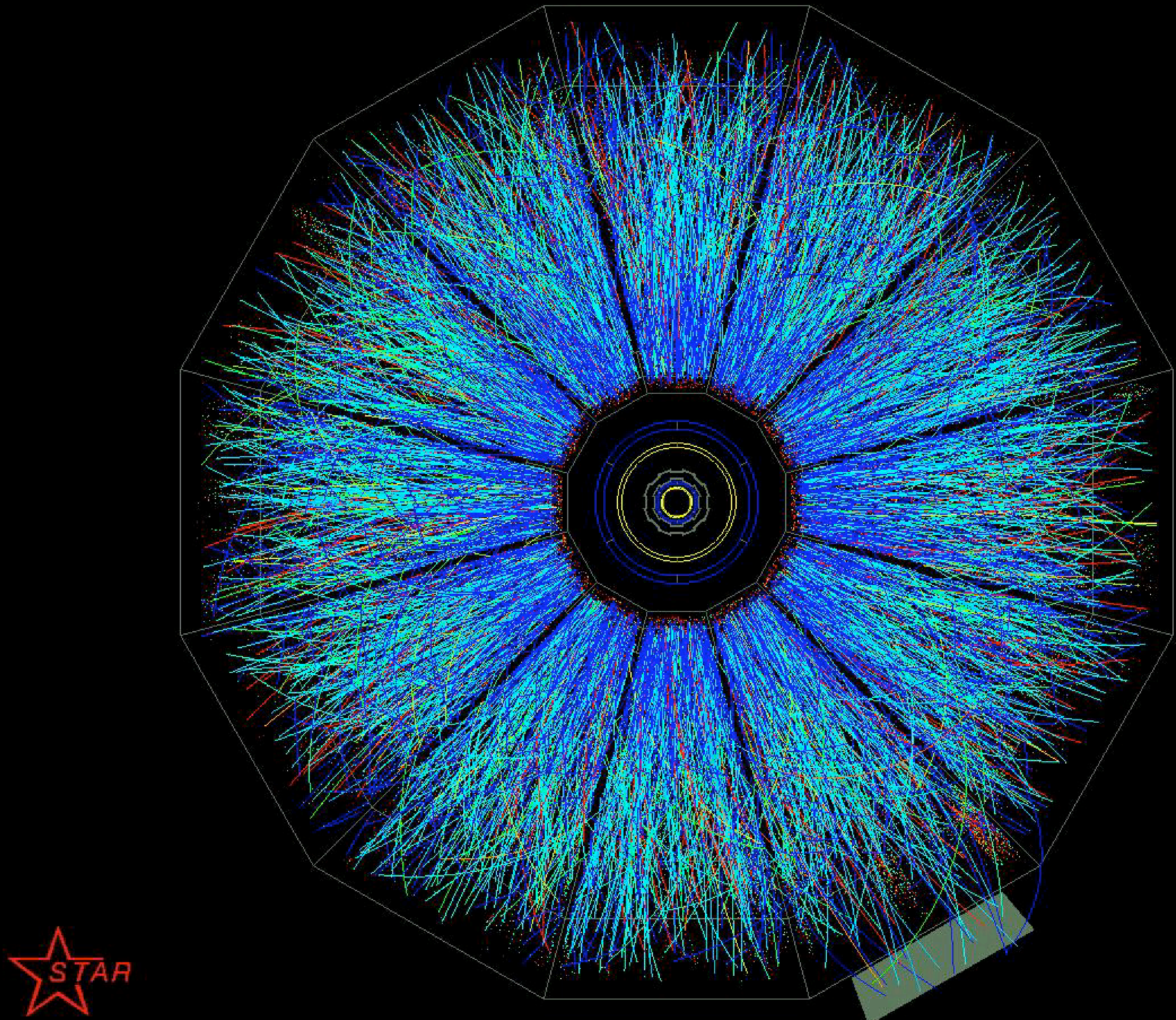


CP-PACS (1998)
GF11 (1993)
experiment
m had [GeV]
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
K
K*
φ
N
Λ
Σ
Ξ
Δ
Σ*
Ξ*
Ω
quenched QCD



Paradigm 3:

The Early Universe Is Simple


STAR

Paradigm 4:

Symmetry Rules

$$\begin{pmatrix} u & u & u \\ d & d & d \end{pmatrix}^{L}_{1/6}$$

$$\begin{pmatrix} \nu \\ e \end{pmatrix}^{L}_{-1/2}$$

$$(u \ u \ u)^{R}_{2/3}$$

$$(d \ d \ d)^{R}_{-1/3}$$

$$(e)^{R}_{-1}$$

$$SU(3) \times SU(2) \times U(1)$$

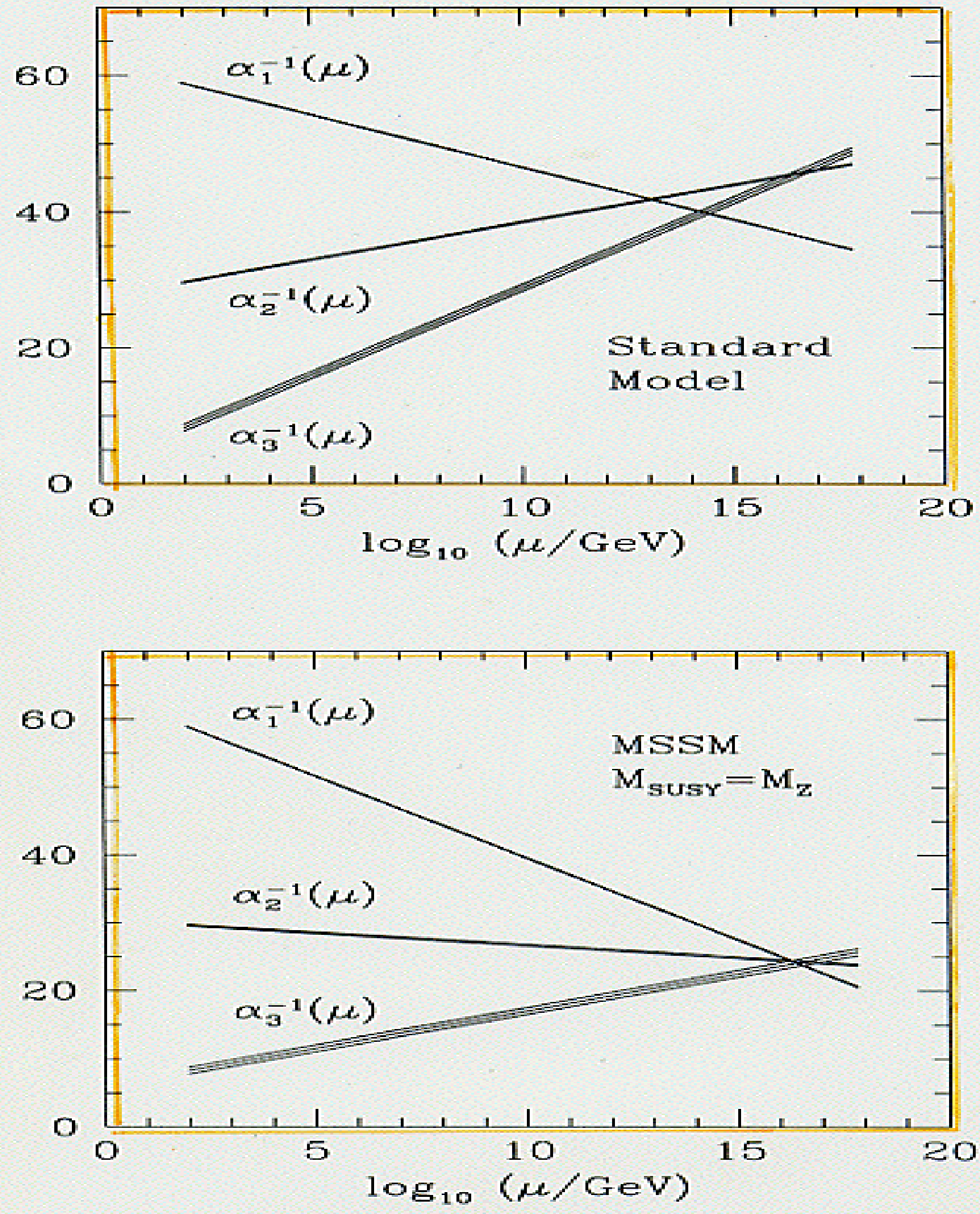
mixed, not unified

No ν^R

	B	R	G	P	O
u	+	−	−	+	−
u	−	+	−	+	−
u	−	−	+	+	−
d	+	−	−	−	+
d	−	+	−	−	+
d	−	−	+	−	+
ν	+	+	+	−	+
e	+	+	+	+	−
u^c	−	+	+	−	−
u^c	+	−	+	−	−
u^c	+	+	−	−	−
d^c	−	+	+	+	+
d^c	+	−	+	+	+
d^c	+	+	−	+	+
e^c	−	−	−	+	+
ν^c	−	−	−	−	−

$$Y = \frac{1}{4}(P+O) - \frac{1}{6}(B+R+G)$$

Unification of gauge couplings...



Gravity fits too! (roughly)

Frontiers of Symmetry

Unification → Proton Decay, Supersymmetry

Supersymmetry → World x2, Dark Matter

QCD T-protection → Axions, Dark Matter

Gauge Symmetry Breaking → Higgs sector

The Greatest Lesson

If we work to understand, then we can understand.

Credits

hadron tables: Particle Data Group

jet event: L3 collaboration

running coupling plot: S. Bethke

pion fields: G. Kilcup

QCD “lava lamp”: D. Leinweber

little bang: STAR collaboration

technical assistance: C. Suggs

The Discovery of Asymptotic Freedom
&
The Emergence of QCD

David Gross

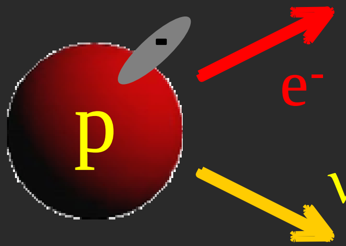
Nobel Lecture
December 8, 2004

The Weak and the Strong

The forces operating in the nucleus are of two kinds:

WEAK INTERACTIONS

Responsible for radioactivity



Fermi theory

STRONG INTERACTIONS

Responsible for holding the nucleus together

Meson theory

QUANTUM FIELD THEORY

The Strong Interactions Were Especially Intractable

- Which particles are elementary:
  p,n,π,..K, Σ ,Λ ,ρ...
- What are the Dynamics?
- How to calculate?

DYSON: *“The correct theory will not be found in the next hundred years. “ (1960)*

A Revolution Was Needed

The Attack on Field Theory

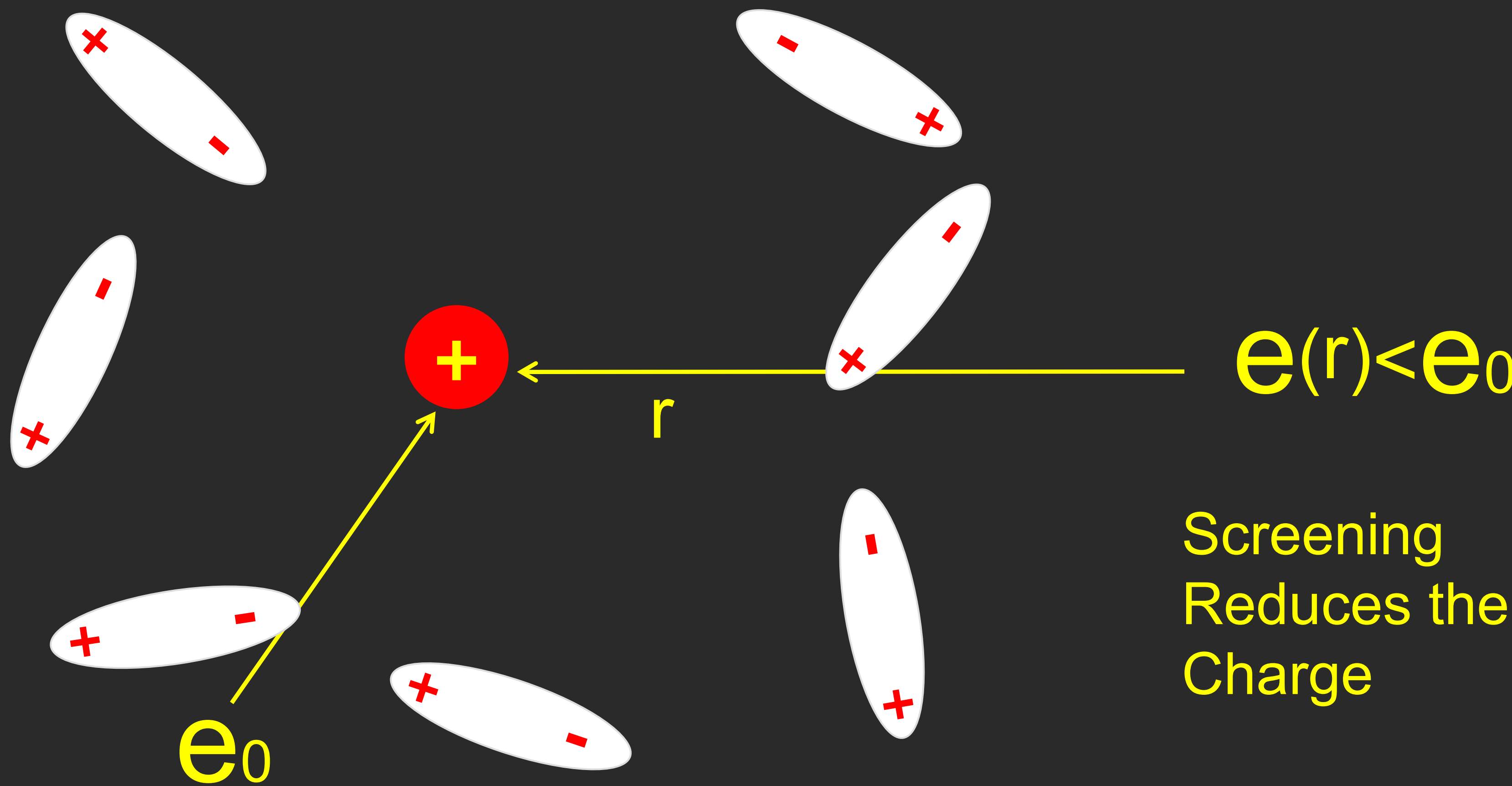
NUCLEAR DEMOCRACY

All hadrons are equally fundamental

BOOTSTRAP THEORY

General principles determine a unique S-Matrix

Screening in Q.E.D.



Screening Reduces the Charge

$$\beta(e) \equiv -\frac{d \ln e(r)}{d \ln(r)} > 0$$

FORCE IS STRONGER AT SHORT DISTANCES

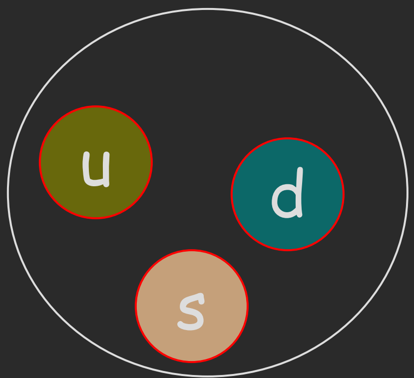
We reach the conclusion that within the limits of formal electrodynamics a point interaction is equivalent, for any intensity whatever, to no interaction at all. *We are driven to the conclusion that the Hamiltonian method for strong interaction is dead* and must be buried, although of course with deserved honor.

Landau (1960)

Patterns & Symmetries

Hadrons looked as if they were made of QUARKS

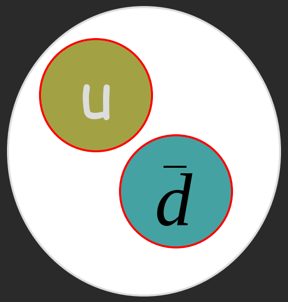
Gell-Mann & Zweig '64



baryons

3 DIFFERENT FLAVORS:

up, down & strange



mesons

And each quark came in 3 identical colors:

BUT QUARKS COULD NOT BE SEEN THEREFORE THEY WERE UNREAL MATHEMATICAL ENTITIES

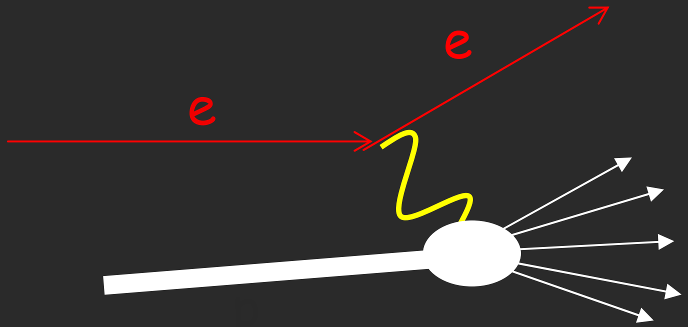
Han-Nambu & Greenberg '64

Berkeley: S-Matrix Theory

I derived (with Callan) some relations-sum rules abstracted From the quark-gluon model.

Harvard: Algebra of Currents

These could be tested in deep-inelastic lepton-hadron scattering experiments (SLAC 1968)



$$\text{Spin} = \frac{1}{2} \Rightarrow \qquad R = \frac{\sigma_L}{\sigma_T} \rightarrow 0$$

1. Hadrons were made of point like constituents.

2. The charged constituents were quarks. Quarks are real.

The Plan:

1. Scaling → Asymptotic Freedom

C. Callan & D.G., 1973

2. There are no Asymptotically Free Field Theories

S. Coleman & D.G., 1973

The one exception:
Non-Abelian Gauge Theories.

With F. Wilczek we determined to close the last hole in the argument (Non-Abelian gauge theories)

We Found

$$\beta(g) = -\frac{g^3}{16\pi^2}\left(\frac{11}{3}N_C - \frac{4}{3}\frac{N_F}{2}\right)$$

Number of colors, quarks

March 1973

Instead of: *No field theory can explain scaling*

There exists a unique field theory that explains scaling

ASYMPTOTICALLY FREE GAUGE THEORY !

Ultraviolet Behavior of Non-Abelian Gauge Theories*

David J. Gross† and Frank Wilczek
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540
(Received 27 April 1973)

It is shown that a wide class of non-Abelian gauge theories have, up to calculable logarithmic corrections, free-field–theory asymptotic behavior. It is suggested that Bjorken scaling may be obtained from strong-interaction dynamics based on non-Abelian gauge symmetry.

Non-Abelian gauge theories have received much attention recently as a means of constructing unified and renormalizable theories of the weak and electromagnetic interactions.[1] In this note we report on an investigation of the ultraviolet (UV) 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.

The UV behavior of renormalizable field theories can be discussed using the renormalization-group equations,[2,3] which for a theory involving one field (say $g\varphi^4$) are

$$[m\partial/\partial m+\beta(g)\,\partial/\partial g-n\gamma(g)]\Gamma_{\rm asy}{}^{(n)}(g;P_1,\ldots,P_n)=0. \qquad (1)$$

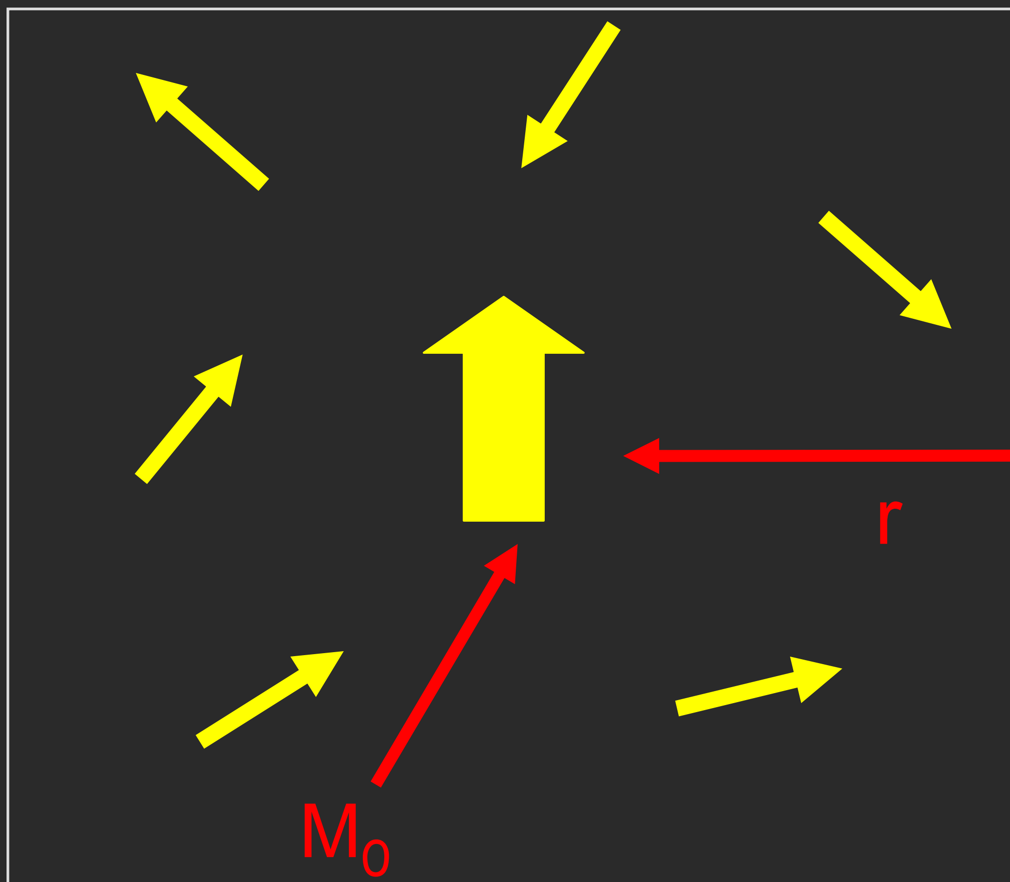
$\Gamma_{\rm asy}{}^{(n)}$ is the asymptotic part of the one-particle–irreducible renormalized n-particle Green's function, $\beta(g)$ and $\gamma(g)$ are finite functions of the renormalized coupling constant g, and m is either the renor-

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 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.

D. J. Gross and F. Wilczek, *Phys. Rev.Lett.* 30, 1343 (1973)

The Anti-screening of QCD



Anti-Screening Increases the Charge.

$$\beta(e) \equiv -\frac{d \ln e(r)}{d \ln(r)} < 0$$

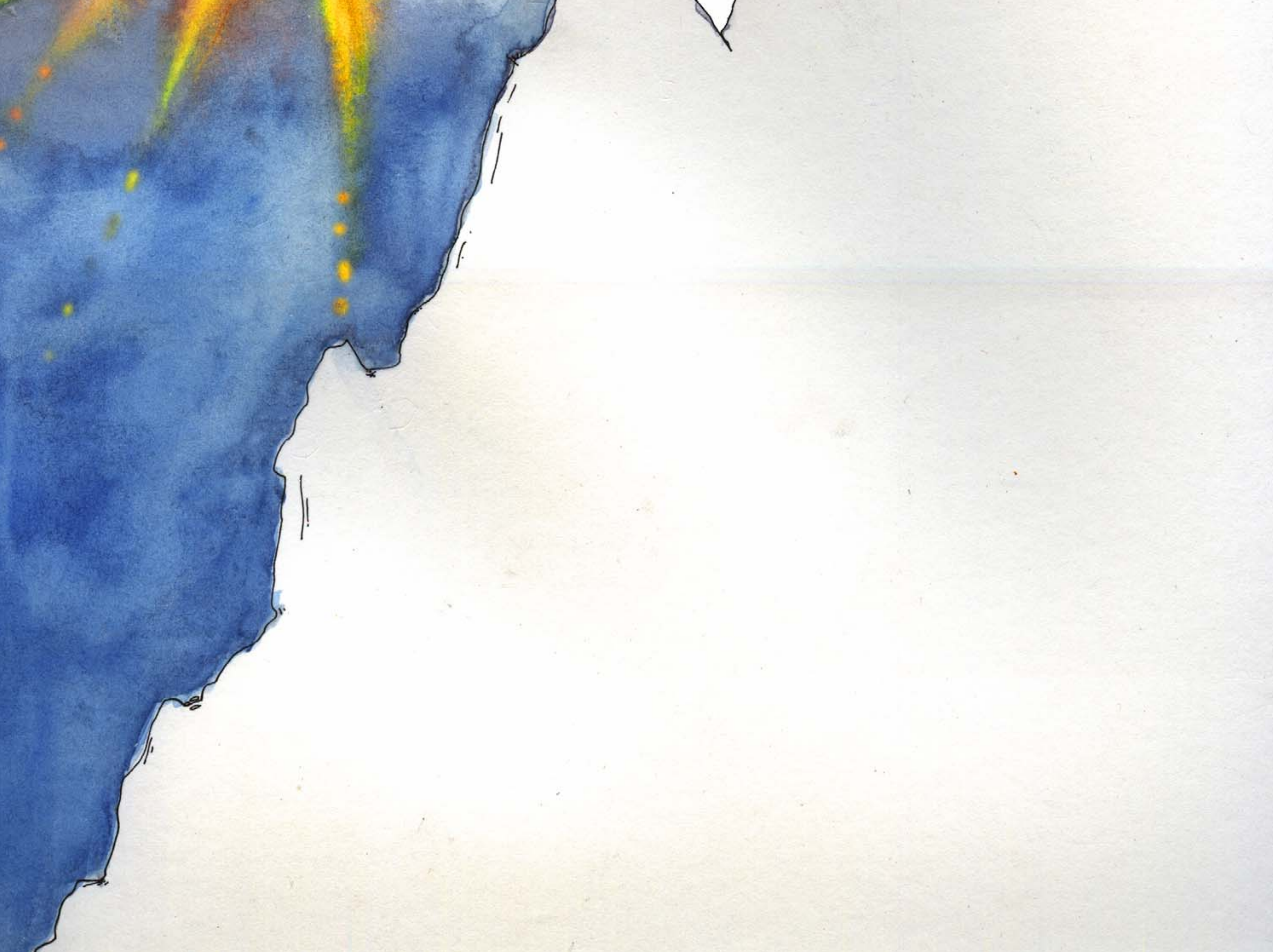
FORCE IS WEAKER
AT SHORT DISTANCES

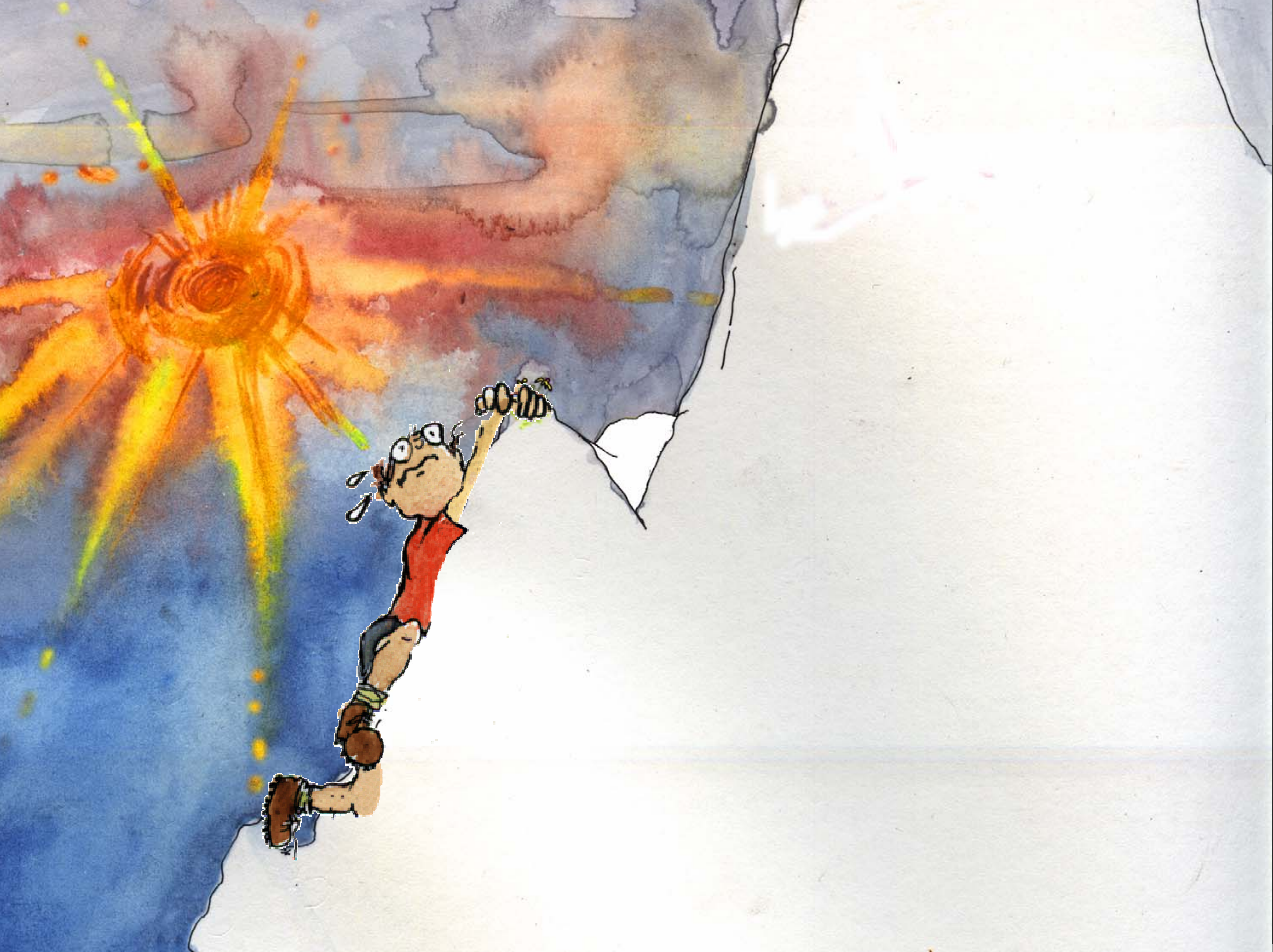
Dynamics ---- Non-Abelian Gauge Theory --- Gluons

Charged Matter --- Exp --- Quarks & 3 Colors

One particularly appealing model is based on three triplets of fermions, with a SU3 color gauge group to provide the strong interactions.

QCD





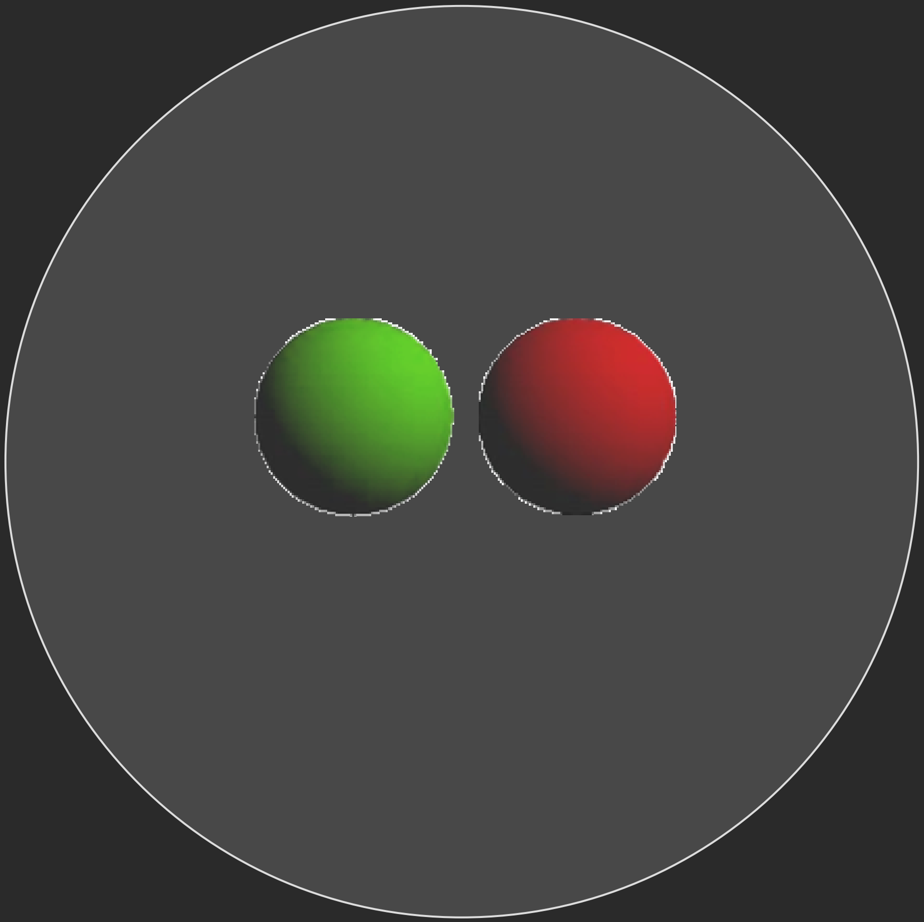
“At first, sight this would appear to be ridiculous since it would imply the existence of massless, strongly coupled vector mesons. However, in asymptotically free theories these naïve expectations might be wrong. There may be little connection between the ‘free’ Lagrangian and the spectrum of states. The infrared behavior of Greens 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.”

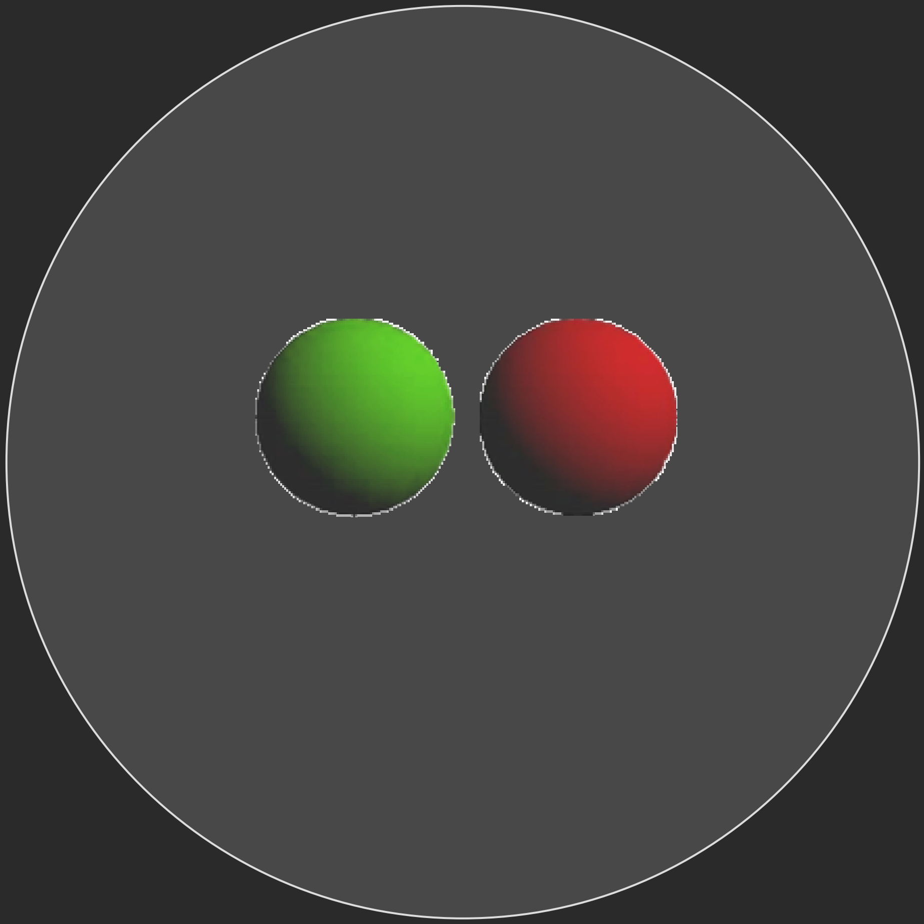
INFRARED SLAVERY → CONFINEMENT

Phys. Rev. D8 30, 3633 (2973)

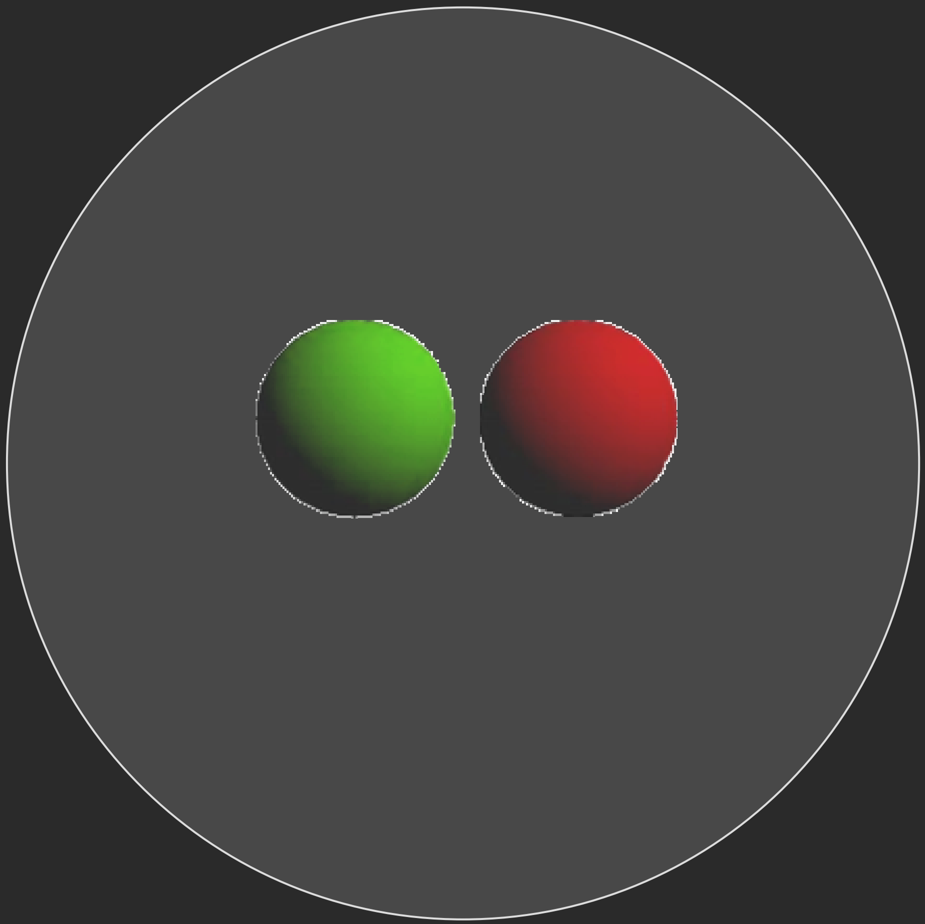
Asymptotic Freedom



Asymptotic Freedom



At short distances, quarks behave freely...



At large distances the quarks are confined.

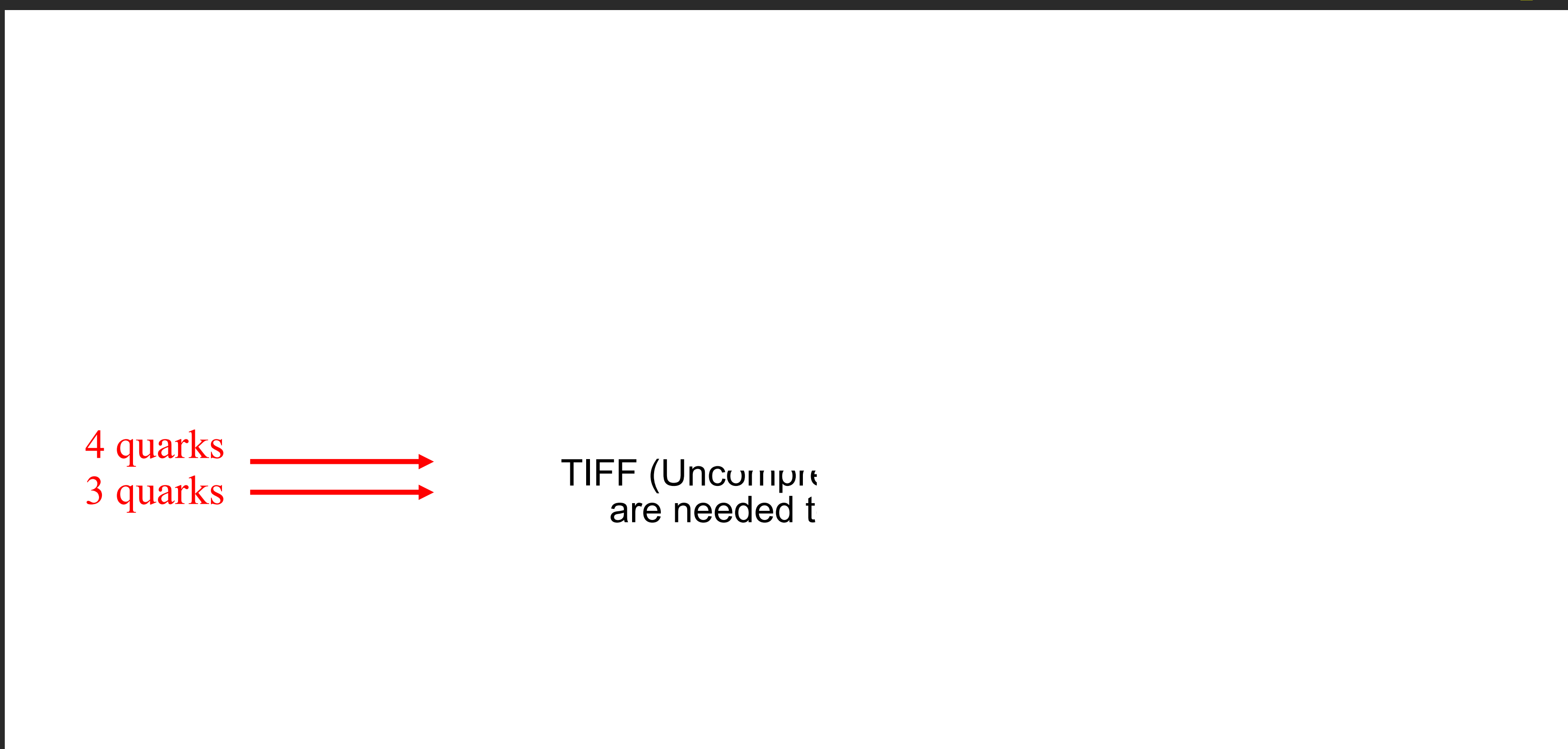
But as the quarks are pulled apart...

Experimental Confirmation

In QCD:

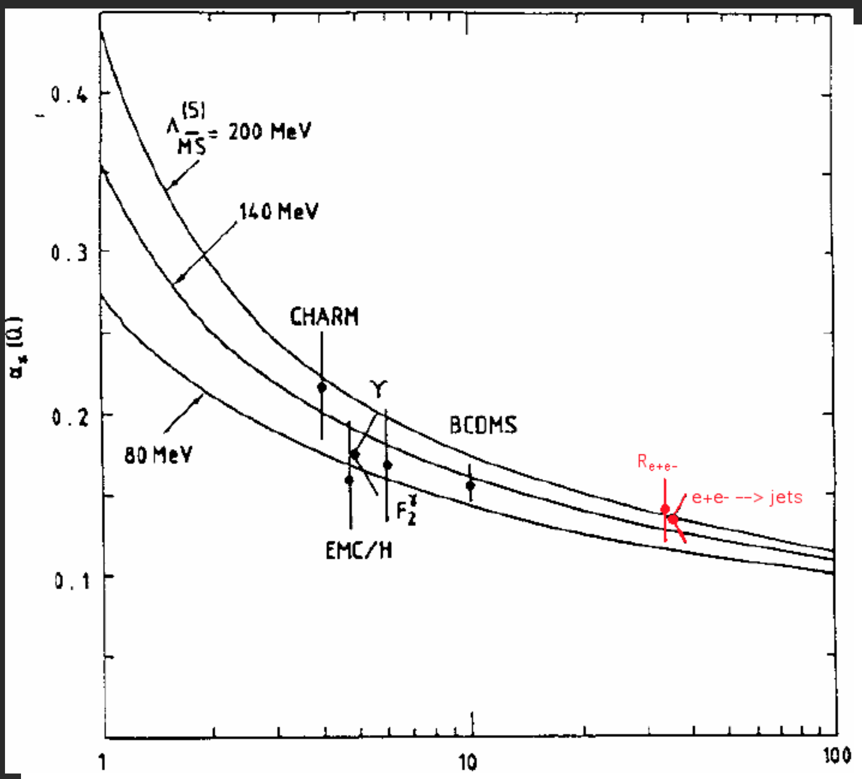
$$R = \frac{\sigma(e^+e^- \rightarrow hadrons)}{\sigma(e^+e^- \rightarrow e^+e^-)} \rightarrow \sum_{QUARKS} Q_i^2$$

Measures the Number and charges of quarks



World summary of α_s

1989	2004



$$\alpha_s(M_z) = 0.110^{+0.006}_{-0.008} \text{ (NLO)}$$

G. Altarelli, Ann. Rev. Nucl. Part. Sci. 39, 1989

Zur Anzeige wird der QuickTime™ Dekompressor „TIFF (LZW)" benötigt.

$$\alpha_s(M_z) = 0.1182 \pm 0.0027 \text{(NNLO)}$$

S. Bethke, hep-ex/0407021

World summary of α_s

Zur Anzeige wird der QuickTime™
Dekompressor „TIFF (LZW)“
benötigt.

S. Bethke,
hep-ex/0407021

Zur Anzeige wird der QuickTime™
Dekompressor „TIFF (LZW)“
benötigt.

Zur Anzeige wird der QuickTime™
Dekompressor „TIFF (LZW)“
benötigt.

World summary of $\alpha_s(M_Z)$

Can all the strong interactions be described by QCD with one single coupling α_s?

YES!

$$\alpha_s(M_Z) = 0.1182 \pm 0.0027$$

S. Bethke, hep-ex/0407021

Implications of Asymptotic Freedom

QCD

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Consistency of Quantum Field Theory

NO DISEASES AT HIGH ENERGY

Asymptotic freedom →
The theory gets simpler at high energy

NO INFINITIES AT SHORT DISTANCES

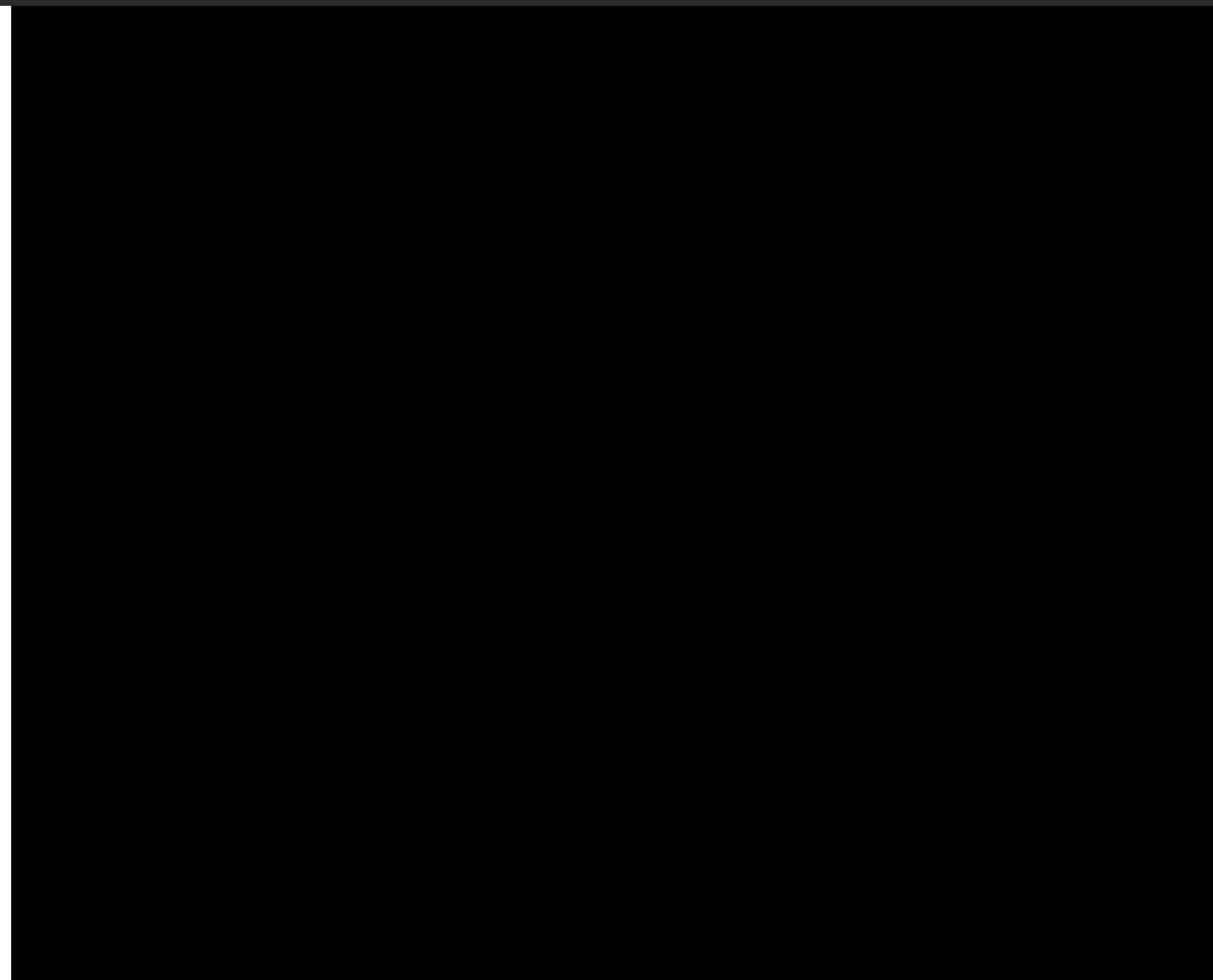
Asymptotic freedom → bare coupling = 0

NO ADJUSTABLE PARAMETERS

All observables are calculable in terms of the dynamically generated mass scale

One can extrapolate QCD to infinite energy and the universe to early times.

Removes the Barrier to Unification



Thank You

NATURE

NOBLE

Thank You

NATURE

NOBLE