

The dilemma of attribution

H. David Politzer*

California Institute of Technology, Pasadena, CA 91125

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 of 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 of 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. In other words, 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 ^N. So far, there's Yang^N, Feynman^N, Glashow^N, Salam^N, Steve Weinberg^N, but not Coleman or Erick Weinberg.)

It turns out there was one brave soul, Tini Veltman^N, who never gave up on Yang–Mills theory, and, with his best-ever grad student, Gerard 't Hooft^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

*E-mail: politzer@theory.caltech.edu.

Adapted from Les Prix Nobel, 2004. © 2004 by the Nobel Foundation

Editor's Note: This article is a version of H. David Politzer's Nobel Lecture, "The Dilemma of Attribution." The 2004 Nobel Prize in Physics was awarded to Drs. Politzer, Frank Wilczek, and David J. Gross for their discovery of asymptotic freedom in the theory of the strong interaction. The Nobel Foundation graciously has granted us permission to reprint this article. The Nobel Lectures provide examples of successful approaches to major scientific problems. However, in recent years, these lectures have rarely been read, perhaps because of the difficulty in obtaining the collections. By reprinting this lecture, we hope to broaden their exposure.

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?” In other words, 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 that 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 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 25 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 to 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, $\pi \rightarrow 2\gamma$, 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

[†]I will add one conjecture to the list, although 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 used 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.

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 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^N, in an experiment that identified a pimple, which Ting^N eponymously titled the J (a reasonable approximation to the relevant Chinese character), on what had been known as Lederman's Shoulder. (That's Leon Lederman^N.) That is to say that Ting's experiment had actually been done earlier by Lederman^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^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 heels 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 Swedish Royal Academy's recognition within just a couple of years of Richter^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 2 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 nonissues, 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—despite 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, were 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. In other words, 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 Academy in identifying noteworthy advances in fundamental physics—and in identifying the particular advance that we celebrate to-

day. 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 corecipients’ 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.

The two books I would recommend first which give excellent accounts of this epoch in particle physics (and more) are *The Second Creation*, by Robert Crease and Charles Mann, and *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 nontechnical audience.

Two marvelous books on 20th century physics for the interested layman which focus more on scientific substance rather than historical process are *The Cosmic Code* and *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 *Les Prix Nobel* 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.