Reflections on a Revolution

John Iliopoulos, reply by Sheldon Lee Glashow

In response to "[The Yang–Mills Model](#)" (Vol. 5, No. 2).

To the editors:

Gauge theories brought about a profound revolution in the way physicists think about the fundamental forces. It is this revolution that is the subject of Sheldon Glashow's essay.

Gauge theories, such as the Yang–Mills model, use two mathematical concepts: group theory, which is the natural language to describe the physical property of symmetry, and differential geometry, which connects in a subtle way symmetry and dynamics.

Although there exist several books, and many more articles, relating historical aspects of these theories,¹ a real history has not yet been written. It may be too early. When a future historian undertakes this task, Glashow's precise, documented, and authoritative essay will prove invaluable. A large part is written in the first person, since the author is one of the main actors in this fantastic play. I cannot think of any nontrivial comments, let alone any criticism, so I will add my views on the subject trying to avoid duplication.

Most revolutions in physics occur when an unexpected experimental result contradicts current theoretical beliefs. But the revolution that brought geometry into physics had a theoretical, and I would say aesthetic, motivation. It came from a few theorists who tried to go beyond a simple phenomenological model.

Most of the time, these theorists played the leading role. Glashow is one of them. As he explains, an important motivation came from the study of what are called weak interactions. It may sound strange that a revolution in particle physics was initiated by the weakest among the forces,² but this happened often before: weak interactions triggered many such revolutions. Maybe we should meditate on the fundamental significance of tiny effects.

In this letter, I touch on several aspects of this revolution and bring some additional elements of history. These include internal symmetries, gauge theories, and the theory of weak interactions.

Internal Symmetries

As Glashow points out, particle physicists distinguish between space-time and internal symmetry transformations. The first change the point of space and time, leaving the fundamental equations unchanged. The second do not affect the space-time point but transform the dynamic variables among themselves. This fundamentally new concept was introduced by Werner Heisenberg in 1932, the year the neutron was discovered, but the real history is more complicated.³ Heisenberg's 1932 papers are an incredible mixture of the old and the new. For many people at that time, the neutron was a new bound state of a proton and an electron, like a small hydrogen atom. Heisenberg does not reject this idea. Although for his work he considers the neutron as a spin one-half Dirac fermion, something incompatible with a proton–electron bound state, he notes that "under suitable circumstances [the neutron] can break up into a proton and an electron in which case the conservation laws of energy and momentum probably do not apply."⁴ In the β -decay controversy that opposed Wolfgang Pauli, who postulated the existence of a neutrino, to Niels Bohr, who advocated non-conservation of energy and momentum, Heisenberg does not take any clear stand, but he sides more with his master Bohr than with his friend Pauli: "The admittedly hypothetical validity of Fermi statistics for neutrons as well as the failure of the energy law in β -decay proves the inapplicability of present quantum mechanics to the structure of the neutron." In fact, Heisenberg's fundamental contribution should be appreciated not *despite* these shortcomings, but precisely *because* of them. It should be remembered that in 1932, experimental data on nuclear forces were almost entirely absent. Heisenberg had to guess the values of the nuclear attractive forces between nucleon pairs by using a strange analogy with molecular forces. He postulated a $p - n$ and an $n - n$ nuclear force, but not a $p - p$ one, so his theory was not really isospin invariant. Nevertheless, he made the conceptual step to describe protons and neutrons together and to introduce the idea of rotations in an abstract space that would interchange a proton and a neutron.⁵ In the following years, three important developments allowed

Heisenberg's initial suggestion to become a complete isospin invariant theory.

The first, and probably the most important, was the progress in experimental techniques, which brought more detailed and more precise data. They showed the need for the introduction of a $p - p$ force and confirmed the charge independence of all nuclear forces.

The second was the 1933 formulation by Enrico Fermi of a theoretical model for the amplitude of neutron β -decay. The influence of this work by far exceeds this initial problem and covers subjects even beyond the theory of elementary particles. Since that time, quantum field theory has become the universal language and one of the cornerstones of modern theoretical physics. To every particle corresponds a quantum field that describes the excitations that physicists call particles.

The third was Hideki Yukawa's introduction of the meson as an intermediary for the nuclear forces. In 1938, Nicholas Kemmer had the idea to incorporate Yukawa's meson into Heisenberg's isospin formalism and write the first fully isospin invariant pion-nucleon interaction. He assumed the existence of three pions with charges $+1$, -1 , and 0 , which he grouped together in an isospin triplet. In 1938, there was evidence for the existence of a charged Yukawa meson, although it was the wrong one, but there was no evidence for a neutral one. Kemmer understood that isospin symmetry required such a neutral meson. In 1950, π^0 was discovered at the Berkeley electron synchrotron by Jack Steinberger, Wolfgang Panofsky, and Jack Steller. Thus π^0 is the first particle whose existence was predicted as a requirement for an internal symmetry and the first to be discovered in an accelerator.⁶

It took six years, as well as the work of many physicists, for Heisenberg's original suggestion of 1932 to become the full isospin symmetry of hadronic physics we know today. It is a remarkably short time, given the revolutionary nature of the idea. The conceptual change has been very important: for the first time, nontrivial internal symmetries have been considered in physics. Heisenberg's isospin space was three-dimensional, and the transformations look like familiar rotations. The concept was subsequently enlarged as new particles were discovered and larger internal symmetry groups were brought into evidence. The space for elementary particle physics became a multi-dimensional manifold, with complicated geometrical and topological properties, and only a subspace of it, the four-dimensional Minkowski space, is directly accessible to our senses.

Gauge Theories and Geometry

The origins of gauge theory can be traced back to classical electrodynamics,⁷ but the importance it has acquired today is due to quantum mechanics. Vladimir Fock in 1926 and Erwin Schrödinger himself were the first to realize that the invariance under local transformations of the phase

of the wave function in the Schrödinger theory implies the introduction of an electromagnetic field.⁸ Naturally, one would expect non-Abelian gauge theories to be constructed following the same principle immediately after isospin symmetry was established in the 1930s. But here history took an unexpected route.

The development of the general theory of relativity offered a new paradigm for a gauge theory. In the next decades, it became the starting point for all studies on theories invariant under local transformations trying to unify gravity and electromagnetism, the only forces known at the time. Theodor Kaluza's attempt, completed by Oskar Klein,⁹ is today often used in supergravity and superstring theories.

Particle physicists date the birth of non-Abelian gauge theories to 1954, with the publication of the fundamental paper by Chen Ning Yang and Robert Mills.¹⁰ The impact of this work on high-energy physics has often been emphasized, but here I want to mention some earlier and little known attempts which, according to present views, have followed a very strange route.

The first is due to Klein. In an obscure conference in 1938, he presented a paper with the title "On the Theory of Charged Fields," in which he attempts to construct an $SU(2)$ gauge theory for the nuclear forces.¹¹ This paper follows an incredibly circuitous road: Klein considers general relativity in a five-dimensional space, he compactifies à la Kaluza-Klein, but he takes the $g_{4\mu}$ components of the metric tensor to be 2×2 matrices. In spite of several technical problems, he finds the correct expression for the field strength tensor of $SU(2)$. He was adding mass terms by hand, and it is not clear whether he worried about the resulting breaking of gauge invariance. It is not known whether this paper has inspired anybody else's work, and Klein himself mentioned it only once in a 1955 conference in Bern.¹²

The second work in the same spirit is due to Pauli, who in 1953, in a letter to Abraham Pais,¹³ developed precisely this approach: the construction of the $SU(2)$ gauge theory as the flat space limit of a compactified higher dimensional theory of general relativity. He was closer to the approach followed today because he considered a six-dimensional theory with the compact space forming an S_2 . He never published this work, and we do not know whether he was aware of Klein's 1938 paper. He had realized that a mass term for the gauge bosons breaks the invariance, and he had an animated argument during a seminar by Yang at the Institute for Advanced Studies in Princeton in 1954.¹⁴ What is certainly surprising is that both Klein and Pauli, fifteen years apart one from the other, decided to construct the $SU(2)$ gauge theory for strong interactions and both chose to follow this totally counterintuitive method. It seems that the fascination exerted by general relativity on this generation of physicists was such that, for many years, local transformations could not be conceived inde-

pendently of general coordinate transformations. Yang and Mills were the first to understand that the gauge theory of an internal symmetry takes place in a fixed background space, which can be chosen to be flat, in which case general relativity plays no role.

I said earlier that the natural mathematical language for gauge theories is differential geometry. I would like to develop this point a bit further while also avoiding technical details as much as possible. For that, I will formulate the theory on a space-time lattice. Lattice field theory is poor man's differential geometry.

Consider, for simplicity, a lattice with hypercubic symmetry. The space-time point x is replaced by a lattice point labeled n , $n = 1, 2, \dots, N$, where N is the number of points of our lattice. A field theory involves the fields $\Phi(x)$ and their derivatives $\partial_\mu \Phi(x)$. The dictionary between these quantities defined in the continuum and the corresponding ones on the lattice is easy to establish (we take the lattice spacing equal to one): $\Phi(x) \Rightarrow \Phi_n$ and $\partial_\mu \Phi(x) \Rightarrow \Phi_n - \Phi_{n+\mu}$, where $n + \mu$ should be understood as a unit vector joining the point n with its nearest neighbor in the direction μ . It is an easy exercise for a student to derive the following results:

1. Ordinary fields Φ_n carry one index n , which means that they live on the lattice points.
2. Invariance under *gauge* transformations implies the introduction of new fields, *the gauge fields*. On the lattice they are denoted by $A_{n, n+\mu}$. They carry two indices n and $n + \mu$, which means that they live on the link from the point n to its nearest neighbor in the direction μ . We say that gauge fields *connect* nearest neighbors, and therefore they live on oriented lattice links. The mathematicians are right when they do not call $A_{n, n+\mu}$ a field, but a connection.
3. New fields imply new interactions; therefore, invariance under gauge transformations implies the introduction of *gauge interactions*.
4. From the geometrical point of view, connecting in a nontrivial way nearest points induces a nontrivial geometry in the fields defined on the lattice. We uncover a deep connection between gauge interactions and the geometry in the field space. This connection between geometry and dynamics is the most profound consequence of gauge theories.¹⁵

Weak Interactions

I will end with a few scattered historical remarks that will complete those mentioned by Glashow.

1. Until the late 1960s, weak interactions were described by Fermi's model. I want to insist that it was phenomenologically very successful. The trouble was theoretical: the results of the calculations

could not be trusted above a certain energy scale, Λ . A naive estimation gives $\Lambda \sim 300$ GeV, which, for physicists in the 1960s, was essentially infinite and nothing to worry about. It was Boris Ioffe and E. P. Shabalin,¹⁶ from the Soviet Union, who first remarked that, in fact, one can do much better. Looking at so-called forbidden processes, such as parity violating effects in strong interactions, or rare decays like $K^0 \rightarrow \mu^+ + \mu^-$, they concluded that the value of Λ was much lower: $\Lambda \sim 3$ GeV. For some people, including myself, this result showed that improving the weak interactions high energy behavior was an urgent question. The first step toward the solution of the problem was taken in 1968 by Claude Bouchiat, Jacques Prentki, and myself,¹⁷ and the complete solution was the introduction of the charm quark in the Glashow–Iliopoulos–Maiani mechanism.¹⁸

2. Glashow mentions a paper he wrote with Murray Gell-Mann in 1961.¹⁹ This paper contains many important points, in addition to those Glashow mentions. First, it correctly identifies the problem related to the absence of $K^0 \rightarrow \mu^+ + \mu^-$ decays, the problem that was solved with the introduction of charm almost ten years later. Second, in this paper, the authors extend the Yang–Mills construction, which was originally done for SU(2), to arbitrary Lie algebras. The well-known result of associating a coupling constant to every simple factor in the algebra appeared for the first time in this paper. Even the seed for a grand unified theory was there. In a footnote they remark that the “remarkable universality of the electric charge would be better understood were the photon not merely a singlet, but a member of a family of vector mesons comprising a simple partially gauge-invariant theory.” This road was followed by Howard Georgi and Glashow in a series of papers which opened a new field of research, that of grand unified theories.²⁰
3. As a last point, I would like to expand on a remark in Glashow's essay. The Yang–Mills model, like any gauge theory, is a theory of currents. They generalize the familiar electromagnetic current. It follows that mathematical consistency requires these currents to be conserved. A seemingly technical point is that, for a quantum field theory, axial currents may not be conserved. In physics jargon, axial currents may have *anomalies*.²¹ For quantum electrodynamics the non-conservation of the axial current can be considered as a curiosity because this current does not play any direct physical role. However, in the electroweak theory, both vector and axial currents are important; the conservation of both is needed. The axial anomaly breaks this conservation and the resulting theory is mathematically inconsistent. The

solution involves a subtle cancellation between the anomalies in the quark and lepton currents.²² It can be expressed as a condition in the electric charges of quarks and leptons in a family: $\sum_i Q_i = 0$,²³ where the sum extends over all fermions in a given family and Q_i is the electric charge of the i th fermion. In other words, *families must be complete*. Thus, the discovery of a new lepton, the tau, implied the existence of two new quarks, the b and the t , a prediction that was verified experimentally. In fact, this condition has a wider application. The Standard Model could have been invented after the Yang–Mills theory was written, much before the discovery of the quarks. At that time, the elementary particles were thought to be the electron and its neutrino, the proton and the neutron. The condition is satisfied. When quarks were discovered, we changed from nucleons to quarks. The condition is again satisfied. If tomorrow we find that leptons or quarks are composite, new building blocks will be required to satisfy this condition again. The important point is that the contribution of a chiral fermion to the anomaly is independent of its mass, so it must be the same no matter which mass scale is used to compute it. Since gauge theories are believed to describe all fundamental interactions, the anomaly cancellation condition plays an important role not only in the framework of the Standard Model, but also in all modern attempts to go beyond, from grand unified theories to superstrings. It is remarkable that this seemingly obscure, higher-order effect dictates, to a certain extent, the structure of the world.

Glashow's essay relates a great achievement of modern theoretical physics. He has been one of its protagonists. It provides material for thought and inspiration. Readers of my generation will wish to add their own ideas, views, and experiences. Young readers may decide to search deeper into gauge theories and hopefully find the way to go beyond. I am sure that this would be Glashow's greatest satisfaction.

John Iliopoulos

Sheldon Lee Glashow replies:

John Iliopoulos provides fascinating backstories to the origins of internal symmetry groups, gauge theories, and our understanding of the weak interactions. My dear friend and erstwhile colleague and collaborator offers neither nontrivial comments about, nor criticism of, my essay. Instead he provides several interesting and valuable historical supplements to my tale. Soon after Ernest Rutherford discovered the proton in 1919, he coined the word "neutron" for a conjectured neutral nuclear constituent

made of an electron bound to a proton. After a near miss by Irène and Frédéric Joliot-Curie, the neutron was discovered by James Chadwick in 1932. Iliopoulos reminds us that many physicists viewed neutrons as collapsed hydrogen atoms, even Werner Heisenberg. Let me add that Rutherford, Niels Bohr, and even Chadwick himself were of that persuasion. Fermi's brilliant 1933 paper outlining his theory of beta decay fails to mention either positrons or antiparticles. Instead he appeals to Paul Dirac's occupied sea of phantom particles to suppress unphysical negative-energy states. The whole truth could not emerge until the discovery of the positron by Carl Anderson later in 1932 and the observation by the Joliot-Curies in 1934 of positron emission accompanying their newly discovered radioactive decay process wherein nuclear protons are transformed into neutrons. Only then could neutrons attain equal status with protons as seemingly elementary particles. I found Iliopoulos's discussion of the attempts by Oskar Klein and Wolfgang Pauli to develop a gauge theory of the strong nuclear force to be utterly fascinating and perhaps even relevant today. It was remarkable to learn how Einstein's general theory led them both astray, as it did Einstein himself. In his discussion of weak interactions, Iliopoulos fails to mention our bold but bootless attempts to make sense of the electroweak theory prior to the brilliant and bountiful efforts of Gerard 't Hooft and Martinus Veltman just one year later.²⁴ Our collaborator in the second of these papers, my then graduate student at Harvard Andrew Chi-Chih Yao, went on to earn a second doctorate in computer science and has since won the Knuth, Pólya, and Turing Prizes for computer science and mathematics. Andrew is now professor and dean at Tsinghua University and a member of both the National and Chinese Academies of Science. I am proud to have had him as a student half a century ago, and today as both a friend and member of the *Inference* Board of Editors.

John Iliopoulos is a CNRS Director of Research Emeritus at the Ecole Normale Supérieure in Paris, where for many years, he was the head of the Theoretical Physics Department.

Sheldon Lee Glashow is a Nobel Laureate, Higgins Professor of Physics, emeritus, at Harvard University, and University Professor, emeritus, at Boston University.

1. See, for example, *The Rise of the Standard Model*, ed. Laurie Brown et al. (New York: Cambridge University Press, 1997), or *The Standard Theory of Particle Physics: Essays to Celebrate CERN's 60th Anniversary*, ed. Luciano Maiani and Luigi Rolandi (Hackensack: World Scientific, 2016).

2. The effects of the gravitational interactions are not measurable in high-energy physics.
3. Strictly speaking, this is not absolutely correct. Already in 1926, people had realized that both the Schrödinger equation and the normalization condition are invariant under a multiplication of the wave function by a constant phase. This is an internal symmetry since the phase transformations do not affect the space-time point x . Nevertheless, they had not been viewed this way and, from the group theory point of view, the phase transformations are trivial.
4. Werner Heisenberg, “Über den Bau der Atomkerne: I (About the Construction of Atomic Nuclei: I),” *Zeitschrift für Physik* 77 (1932): 1–2, doi:10.1007/BF01342433.
5. Heisenberg called it *Platzwechsel*, or position-exchange, interaction.
6. Kemmer is a strange case in the history of particle physics. His expression for the pion–nucleon interaction can be found in every textbook, yet his name is practically never mentioned. As Freeman Dyson remarked, blowing the trumpets for his own success was not in Kemmer’s character.
7. For an excellent historical account, see Olivier Darrigol, *Electrodynamics from Ampère to Einstein* (New York: Oxford University Press, 2000).
8. Vladimir Fock, “On the Invariant Form of the Wave and Motion Equations for a Charged Point Mass,” *Physics-Uspekhi* 53, no. 8 (2010): 839–41, doi:10.3367/ufne.0180.201008h.0874. This is a translation from the original “Über die invariante Form der Wellen- und der Bewegungsgleichungen für einen geladenen Massenpunkt,” *Zeitschrift für Physik* 39, no. 2–3 (1926): 226–32, doi:10.1007/bf01321989.
9. Theodor Kaluza, “Zum Unitätsproblem der Physik (On the Unity Problem of Physics),” *Sitzungsberichte Preußische Akademie der Wissenschaften* (1921): 966–72; Oskar Klein, “Quantentheorie und fünfdimensionale Relativitätstheorie (Quantum Theory and a Five-Dimensional Theory of Relativity),” *Zeitschrift für Physik* 37, no. 12 (1926): 895–906, doi:10.1007/bf01397481.
10. Chen Ning Yang and Robert Mills, “Conservation of Isotopic Spin and Isotopic Gauge Invariance,” *Physical Review* 96 (1954): 191, doi:10.1103/PhysRev.96.191. It seems that similar results were obtained by Robert Shaw in his thesis.
11. Oskar Klein, “Sur la théorie des champs associés à des particules chargées (On the Theory of Fields Associated with Charged Particles),” in *Les Nouvelles théories de la physique*, proceedings from the conference organized in Warsaw, May 30–June 3, 1938 (Paris: Institut International de Coopération Intellectuelle, 1939), 81.
12. Oskar Klein, “Generalization of Einstein’s Theory of Gravitation Considered from the Point of View of Quantum Field Theory,” in *Fünfzig Jahre Relativitäts-theorie, Bern, 11–16 Juli 1955 (Helvetica Physica Acta Supplementum 4)*, ed. André Mercier and Michel Kervaire (Basel: Birkhäuser, 1956), 58–68.
13. Wolfgang Pauli, unpublished. It is summarized in a letter to Abraham Pais dated July 22–25, 1953. See Pais, *Inward Bound* (Oxford: Oxford University Press, 1986), 584.
14. Chen Ning Yang, *Selected Papers 1945–1980 with Commentary* (San Francisco: Freeman, 1983), 525.
15. In the entrance of Plato’s Academy in classical Athens, it was written “Μηδείς ἀγεωμέτρητος εἰσίτω...” which means “Let no one ignorant of geometry enter...”
16. Boris Ioffe and E. P. Shabalin, “Neutral Currents and the Applicability Limit of the Theory of Weak Interactions,” *Yadernaya Fizika* 6 (1967): 828; English translation in *Soviet Journal of Nuclear Physics* 6 (1968): 603. See also Rabin-dra Nath Mohapatra, J. Subba Rao, and Robert Marshak, “ $K_L^0 \rightarrow \mu^+ \mu^-$ Decay, $K_L^0 - K_S^0$ Mass Difference and Weak-Interaction Cutoff,” *Physical Review Letters* 20 (1968): 1,081–84, doi:10.1103/PhysRevLett.20.1081.
17. Claude Bouchiat, John Iliopoulos, and Jacques Prentki, “Divergences of Weak Nonleptonic Amplitudes and the Breaking of $SU_3 \times SU_3$,” *Nuovo Cimento A* 56 (1968): 1,150–54, doi:10.1007/bf02751632; Iliopoulos, “Leading Divergences to All Orders of Weak Interactions,” *Nuovo Cimento A* 62 (1969): 209–28, doi:10.1007/bf02731803.
18. Sheldon Lee Glashow, John Iliopoulos, and Luciano Maiani, “Weak Interactions with Lepton-Hadron Symmetry,” *Physical Review D* 2, no. 7 (1970): 1,285–92, doi:10.1103/physrevd.2.1285.
19. Sheldon Lee Glashow and Murray Gell-Mann, “Gauge Theories of Vector Particles,” *Annals of Physics* 15, no. 3 (1961): 437–60, doi:10.1016/0003-4916(61)90193-2.
20. Howard Georgi and Sheldon Lee Glashow, “Unified Weak and Electromagnetic Interactions without Neutral Currents,” *Physical Review Letters* 28 (1972): 1,494–97, doi:10.1103/physrevlett.28.1494. In this paper, they propose a model in which the only neutral gauge boson is the photon. Even after the discovery of the weak neutral currents, this model has survived as a toy model because it has one interesting feature: the photon is a gauge boson of a simple Lie algebra. In this sense, it is an elementary version of what is called grand-unified theories. These were introduced by the same authors in “Unity of All Elementary-Particle Forces,” *Physical Review Letters* 32, no. 8 (1974): 438–41, doi:10.1103/physrevlett.32.438.
21. Stephen Adler, “Axial-Vector Vertex in Spinor Electrodynamics,” *Physical Review* 177 (1969): 2,426–38, doi:10.1103/physrev.177.2426; John Stewart Bell and Roman Jackiw, “A PCAC Puzzle: $\pi^0 \rightarrow \gamma\gamma$ in the σ -Model,” *Nuovo Cimento A* 60 (1969): 47–61, doi:10.1007/BF02823296.
22. Claude Bouchiat, John Iliopoulos, and Philippe Meyer, “An Anomaly-Free Version of Weinberg’s Model,” *Physics Letters B* 38, no. 7 (1972): 519–23, doi:10.1016/0370-2693(72)90532-1; David J. Gross and Jackiw, “Effects of Anomalies on Quasi-Renormalization Theories,” *Physical Review D* 6 (1972): 477–93, doi:10.1103/physrevd.6.477; Howard Georgi and Sheldon Lee Glashow, “Gauge Theories without Anomalies,” *Physical Review D* 6 (1972): 429, doi:10.1103/physrevd.6.429.
23. Glashow, Iliopoulos, and Maiani, “Weak Interactions with Lepton-Hadron Symmetry.” With this work, we showed that the title was mathematically justified.

24. Sheldon Lee Glashow and John Iliopoulos, “[Divergences of Massive Yang-Mills Theories](#),” *Physical Review D* 3, no. 4 (1971): 1,043–45, doi:10.1103/physrevd.3.1043; and Sheldon Lee Glashow, John Iliopoulos, and Andrew Yao, “[Divergences of Massive Yang-Mills Theories: Higher Groups](#),” *Physical Review D* 4, no. 6 (1971): 1,918–19, doi:10.1103/physrevd.4.1918. Gerard 't Hooft and Martinus Veltman, “[Regularization and Renormalization of Gauge Fields](#),” *Nuclear Physics B* 44, no. 1 (1972): 189–213, doi:10.1016/0550-3213(72)90279-9.