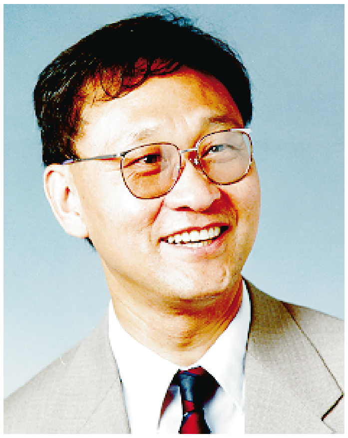

REFLECTIONS ON THE STANDARD MODEL

Tian Yu Cao

Department of Philosophy, Boston University

I. In this talk, I am going to give a brief account of the genesis of the standard model. Then, after giving some assessment of its present situation, I will end with a few words about its future.

The term standard model was first coined by Abraham Pais and Sam Treiman in 1975, with a reference to the electroweak theory with four quarks. In its later usage, as all of you are familiar with, it refers to a system of particles consisting of a six-quark picture of hadrons together with six leptons, with its dynamics described by the electroweak theory and QCD.

As to its genesis, there is a standard story which is widely circulated, and goes as follows. At first, in 1967-68, there was a Weinberg-Salam model of unified electroweak theory, in which the Higgs mechanism was in the first time incorporated into an $SU(2) \times U(1)$ Yang-Mills theory. As a gauge invariant theory, its renormalizability was assured but not explicitly proven, because by choosing a unitary gauge, which made a great physical sense, the inventors of the model found it difficult to give an explicit proof of its renormalizabilty. Then came a young and smart Dutch graduate student Gerard ‘t Hooft, who invented a renormalizable gauge, and with this technical innovation, a renormalizable unified electroweak theory was available. Then, by hindsight, Glashow’s 1961 model was recognized as a precursor to the Weinberg-Salam model. The award of the Nobel prize in 1979 was in accordance with this story, and also endorsed, reinforced the story, and elevated it to a standard story for the genesis of the standard model. As to the QCD part of the model, the story goes, it was the result of a renormalization group calculation by David Gross and Frank Wilcsek in 1973, which led to the discovery of asymptotic freedom of the quark-gluon system, thus a perturbatively renormalizable theory for the system, named QCD, was available. Will Gross and Wilschek get a Nobel prize for their contribution to QCD? I don’t know. Perhaps they will get; perhaps not. It depends, as all of you know, on very complicated politics.

I am not happy with this standard story. The reason for this unhappiness is that it trivializes the intellectual history of the genesis of the standard model, and gives a wrong impression that there was a smooth evolution from the Weinberg-Salam model to the standard model with some technical help from t'Hooft, Gross and Wilczek. I even wish to make a stronger claim that this standard story in fact has distored the real history of contemporary physics, because there was no such smoothness in the history. And in fact, the sudden occurrence of the possibility, in 1971, for a unified model of physical interactions, which would be phenomenologically viable and at the same time conceptually and mathematically consistent, came as a suprise to many physicists, Weinberg, Salam and Glashow included. Some of you may still remember that in the 1960s, the dominant framework for high energy physics was not the Yang-Mills theory, not even quantum field theory, but S-matrix theory and current algebra. It is not too difficult to understand that there was no Royal road leading from S-matrix theory and current algebra to the standard mode. Then what was the reason for the standard model to arise from such an unfavorable theoretical context? Detailed answer to this question can be found in my recently published book by Cambridge University Press, titled CONCEPTUAL DEVELOPMENTS OF 20TH CENTURY FIELD THEORIES. Here I just want to give you a very brief account.

From an intellectual historian's perspective, the rise of the standard model was the result of three lines of development: (i) the establishment of a conceptual framework, (ii) a proof of its consistency, and (iii) model building. The general framework within which the standard model was built was reductionist or atomistic in nature. From the atomistic perspective, all phenomena as appearance should be explained by the ultimate ingredients of the world, the deep and true relity. Although arguably it is the dominant theme and fashion in high energy physics, and is built into core commitment of quantum field theory, with the exceptions of S-matrix theory, current algebra and effective theories, it was not the case for 19th century physics: the substantial development of mechanics, thermodynamics and electromagnetism in the 19th century had little to do with atomism, and the revival of atomism in 20th century physics was stimulated by the successes of statistical mechanics and mainly by the success of quantum physics.

Then what are the elementary ingredients of the standard model? They are not the elementary particles in Wigner's sense, appearing in a unitary representation of Poincare group, such as nucleons and mesons, but quarks (parts of hadrons with baryon number 1/3, spin 1/2, and fractional electric charge) and leptons. The quark model of hadrons originally suggested by Gell-Mann and Zweig with three species, later called flavors, forming a fundamental basis of an $SU(3)$ group as a way of incorporating the broken $SU(3)$ symmetry among hadrons, allows quarks, together with leptons, to be taken as the basic ingredients of the microstructure of the physical world. As mental constructions, quarks were used to construct workable models: first came currents and dual resonance model, then, combined with the idea of non-abelian gauge coupling, or Yang-Mills coupling, which served to fix the dynamics of quarks and leptons, came the Weinberg-Salam model.

Second, in order to have some confidence in the consistency of the Yang-Mills theory, a proof of its renormalizability was theoretically and also psychologically absolutely necessary. At least that was the dominant opinion of the 1960s. Third, in order to build a phenomenologically viable model, physicists needed mechanisms for short range behaviour in weak and strong interactions within a Yang-Mills framework, and also within a perturbative framework, because that was the only framework physicists felt comfortable with.

Now let us turn to the chronology of the development along these three lines. Soon after Yang and Mills proposed in 1954 a non-abelian gauge theory for the strong interactions, came several attempts to building models for fundamental interactions within this framework, which was relatively easier than the proof of its renormalizability but quite fascinating. Early attempts at building models appeared from the mid 1950s by Utilyama, Schwinger, Bludman, Salam and Ward, and in the early 1960s by Sakurai, Glashow and Salam and Ward. After the advent of the quark model and Higgs mechanism in 1964, there appeared, among others, models proposed by Weinberg and Salam. As far as the empirical content is concerned, Glashow's 1961 model contains all essentials of a unified electroweak theory, and later refinements with the incorporation of the Higgs mechanism and the proof of its renormalizability only adds more justifications for its consistency.

The concept of Yang-Mills coupling, which itself as an analogical extension of the minimal coupling well-known in QED, though attractive, was empty were it not enriched, complemented and supported by a net of concepts. Among those complementary concepts we find various concepts of symmetry breaking: the concept of spontaneous breaking developed by Heisenberg, Nambu, Goldstone, Anderson, and many others, and the concept of anomalous breaking by Adler, Bell and Jackiw, and many others. These concepts played an indispensable role in establishing the consistency of the framework and in model building. Most importantly, the anomalous breakdown of scale invariance provided a justification to the new version of the concept of renormalization group, which was developed by Wilson, Callan and Symanzik around 1970, on the basis of Gell-Mann and Low's work of 1954. This concept assumed a specific type of causal connections between the structure of physical interaction at different energy scales, without which no idea of running couplings, and hence no idea of asymptotic freedom would be possible, nor would a rigorous proof of renormalizability based on the fixed point solution of Wilson flow be possible either.

The crucial steps in establishing the physical reality of the quark model were taken in 1969 when deep inelastic scattering experiments were performed at SLAC to probe the short distance structure of hadrons. The observed Bjorken scaling suggested that hadrons consisted of free point-like constituents or partons, some of which, when the experimental data were analyzed in terms of the operator product of currents, turned out to be charged and have baryon number and spin 1/2. That is, they looked like quarks. Further data also showed that they were consistent with the electric charge assignment to quarks. These preliminary but crucial developments certainly had convinced some pioneering physicists of the reality of quarks, and encouraged them to use the quark model for conceptualizing the subatomic world. Otherwise, nobody, not even Gell-Mann himself, would take the quark model seriously as an ontological basis for theorizing the physical world.

All these were fascinating developments. But there was no standard model. The Weinberg-Salam model was not even noticed by the physics community before 1971. Feynman's parton model prompted by the SLAC experiments contributed to the rise of QCD, but as a phenomenological model, it did not take gauge principle seriously. Theoretically speaking, all these models were quite vulnerable. No consistent quantization of the models was given. Without an indication of proper Feynman rules, speculations about renormalizability of whether it would be spoiled by spontaneous symmetry breaking were conceptually empty.

More specifically, as far as the massive vector theory is concerned, the introduction of the Higgs mechanism was an important step in the right direction, although not a decisive one, in establishing the renormalizability of the model. It was through Feynman in 1963, DeWitt from 1964 to 1967, Faddeev and Popov in 1967, to Veltman and 't Hooft from 1968 to 1972, to name just a few most crucial figures. They quantized the theory in a consistent way with the introduction of a complex system of non-physical degrees of freedom required by accepted physical principles. They derived Feynman rules and Ward Identities, invented the renormalizable gauges, and proved unitarity. Finally, they invented a gauge invariant regularization scheme. Without these investigations and achievements, no proof of the renormalizability of non-abelian gauge theories would be possible, and all the convictions and conjectures of an a priori kind, based solely on the symmetry argument and naive power counting argument, would be groundless and empty.

Many physicists have held the opinion, and I agree with them, that the great change came at the Amsterdam conference organized by Martinus Veltman in June 1971, when 't Hooft's proof of the renormalizability of massless Yang-Mills theory was presented publicly. However, it has often been forgotten or ignored that this proof was based on Veltman's earlier work, employed the ideas, techniques and tools developed by Veltman, and was completed under Veltman's supervision, and its correctness was checked carefully by Veltman before the result went to public. With a consistent framework at hand, 't Hooft also proposed three models with great ease. One of 't Hooft's models was identical to the forgotten one, proposed previously by Weinberg and Salam, which, apart from the Higgs mechanism, was empirically equivalent to the model proposed by Glashow in 1961. This was a turning point. This was a critical point for the rise of the standard model. It was critical because it signaled a phase transition in the intellectual climate of the high energy physics community, thereafter physicists have gained their confidence in the consistency of quantum field theory in general and of Yang-Mills theory in particular. Within this changed atmosphere, intensive investigations on the short distance behaviour of Yang-Mills theory carried out by Symanzik, 't Hooft, Parisi, Callan, Gross and Wilczek, soon produced decisive result of asymptotic freedom, which, combined with

Nambu's idea of taking color symmetry as a gauge symmetry for the quark-gluon dynamics, ushered in QCD immediately. And we have the standard model now.

The point I wish to make is that without a consistent conceptual framework, a quark model combined with a dynamics dictated by gauge couplings, there would be no standard model, and without a proof of the renormalizability of the non-abelian gauge theory presented by Veltman and 't Hooft in a psychologically convincing way, there would be no confidence in having a consistent framework. Yet the great pity is that the great contribution by Veltman and 't Hooft to the rise of the standard model has not been properly appreciated. Of course, this contribution itself was prepared by other pioneering works of Lee and Yang, Feynman, DeWitt, and many others. Thus, it was not an achievement of a few persons, but an achievement of the physics community, which was guided by the general agenda and internal logic of quantum field theory. So much for the genesis of the standard model.

II. Then what is my assessment of the standard model? First, it is very successful. Of course it doesn't say too much. Everybody knows that experimentally it is very successful, W-bosons, top quark, etc. A novelty of its success is that it has created a new world, a man-made world, and it has pushed physics from a world given to a world made. You cannot find a W-boson in the natural world. You have to create it in the laboratory first, and then to detect. And this amazing power of creating a predictable and controllable new world is given by the standard model.

Second, it has provided a framework that has shaped the patterns for theoretical and experimental discoveries in particle physics. It has also offered or rather reinforced a distinctive language, a language of reduction and unification, for the general public to theorize about the world. No one would dispute that our contemporary conception of the world, from the ultimate constitution of matter to the laws of nature to the evolution of the universe, is largely shaped by the standard model.

Third, from the perspective of practicing physicists, however, the standard model, although very successful in the new area it created, has not gone very far beyond its early success in providing explanation for the old area, for example, the strong interactions among hadrons, for which it was originally designed and hoped. More specifically, it is felt that many numerical results in nuclear physics can be calculated and predicted without any information at all about the internal working of the hadrons, thus having inherited very little of importance from QCD.

Fourth, the standard model are facing a set of difficult questions with no satisfactory answers in sight. Here I am not referring to the so-called meta-questions, such as "why there are three and only three generations of fermions?" "why there are vacuum fluctuations?" "why the renormalization group transformations are so smooth and structureless?" which are of course beyond its reach and are unfair to raise in the first place. What are fair to raise to the standard model are a set of intrinsic questions, such as the vertical question of explaining hadrons in terms of quarks and gluons, and the horizontal question of incorporating gravity into the model. Unfortunately, the standard model has not offered enough theoretical resource to attack these legitimate intrinsic questions.

Looking at the situation from a broader perspective, I would like to claim that conceptually speaking, the standard model, after its early success, has entered a long period of stagnation or even crisis in the last two decades, and there is no end or way out in sight. As a result or as a manifestation of this stagnation or crisis, many aspiring physicists have felt that the intellectual excitement of physics research nowadays lies, not in the standard model, but elsewhere, in strings, in black holes or in cosmology or in non-linear dynamics, chous theory, etc.

The reason for this frustration with the standard model is quite complicated. Conceptually, the unification of electroweak with strong interactions has been attacked without success, let alone the quantization of gravity and its unification with other interactions. The explanation of pion-nucleon interactions by QCD seems almost unattainable. Even the self-consistency of the standard model itself seems also to be in a dubious situation. Here what I have in mind are the difficulties related with such issues as quark confinement in 4-dimensional spacetime, Higgs physics, rigorous proof of renormalizability based on the fixed point of renormalization group, etc. In a theory such as the standard model which on the one hand inflates the notion of elementary, and on the other relegates its basic constituents, such as quarks, gluons and Higgs particles, to unobservables, one cannot help but seriously doubt whether in fact the concept of a fundamental entity has any physically objective meaning, and whether the goal of identifying it will ever be reached.

The sense of the failure of the standard model in its reductionist pursuit is further deepened by some important developments dictated by the inner logic of the standard model. What I refer to here are the concepts of symmetry breakings, renormalization group and, most importantly, decoupling. The decoupling theorem rejects, first, the attempt to give causal connections between different levels (universal significance) and, second, the stipulation of their direct relevance to scientific inquiry. It rejects the suggestion that it is possible simply by means of these causal connections to infer the complexity and the novelty that emerge at the lower energy scales, from the simplicity at higher energy scales, without any empirical input.

As I have argued elsewhere, the new picture of the physical world suggested by these concepts is a hierarchy layered into quasi-autonomous domains, separated by mass scales associated with spontaneous symmetry breaking. Connections between different layers exist and can be mathematically expressed by renormalization group equations. These connections manifest themselves most conspicuously in the renormalization effects of high energy processes low phenomena. Yet the ontology and dynamics of each layer, according to the decoupling theorem, are quasi-stable, almost immune to whatever happens at the other layers. Thus the name of quasi-autonomous domains. Such a world picture, in contrast with the picture of a reducible hierarchy, supports the existence of objective emergence, which set an intrinsic limit to the reductionist strategy. Thus when we look at some of the conceptual difficulties mentioned above, we feel that they are unlikely to be normal puzzles that can be solved by the existing methodology. Rather, they may have revealed a deep crisis in the very idea of reductionism which underlies the standard model.

That is, the knowledge acquired at the fundamental level may not be relevant to other levels of the physical world, let alone other spheres of human activities.

Epistemologically speaking, reduction is of high value in explanation; it also intrigues us to probe deeper level of the world, which helps us to have a partially unified description of physical phenomena; but it performed very poor in terms of economy or simplification, because the deeper one digs, the more one discovers, not simplicity, but greater complexity, with no end in sight to the richness and complexity of the world. Ontologically speaking, reduction never succeeded. A typical example for the failure of ontological reductionism is the fact that there is no logically tight and convincing reduction of hadrons to quarks and gluons without involving endless complications about the structure of QCD vacuum and its fluctuations, mass gap, renormalization group transformation, duality of strong and weak coupling phases, and many other things in the endless list of difficulties and challenges. Instead of reduction, the only thing we can say is that quarks and gluons are parts of hadrons. But as Aristotele had already realized, the whole is more than the sum of its parts. And this gives a death sentence to the ontological reductionism.

These days many people are talking about high energy physics in crisis. Generally speaking, a crisis of a subject or an institute is a process of delegitimation fuelled by the demand for legitimating itself. A crisis of a subject in its deepest sense is an internal erosion of its legitimacy rather than its being situated under unfovarable external pressures. Thus a deep sense of crisis within a community is induced only by failure to meet the expectations which are raised and firmly established by its previous successes, and determined by the value system shared by the community. The sense of crisis is further deepend by the shaking of conviction in the strategy that is taken by the community as essential and crucial for attaining the goals set by the community. In the case of high energy physics, what is at stake are the principles of unification and reduction. The expectation raised by the previous successes of the standard model is to have a unified theory of everything which is relevant to various phenomena, so that the resource-expensive researches in high energy physics can be somewhat justified.

The cognitive failure of the reductionist strategy to meet the expectation of unification as we have mentioned just moments ago have induced social failure of the high energy physics community in its negotiation with the society at large: The argument of irrelevance is (a fortiori) applicable to other areas of human activities. Thus a tension has appeared, and becomes increasingly exacerbated these days, between the high energy physics community with its futile reductionist passion and its sponsor, the government and the society at large, whose dominant conception of rationality is instrumental in nature.

Thus a conflict between the reductionist rationality cherished by the high energy physics community and the instrumental rationality held by the society (looms large, and manifests itself in a social conflict: the physics community finds itself facing unexperienced funding difficulty, declining social status and the shrinkage of job opportunities. These severe social constraints demand a deep and painful structural readjustment of the community, which has an alarming effect upon its size and vitality. Together with the restructing of the community is a change of the research agenda, a change from aiming at the fundamental theory to effective theories.

The change can take two different directions. On the one hand, the community can tactically adapt itself to the pressures stemming from the sponsoring society while maintaining the old conceptual framework, hoping that, with the help of more sophisticated mathematical devices, such as conformal invariance, index theorem and phase cell localization; or by appealing to novel physical ideas, such as those of string, compactification of higher dimensional spacetime and various versions of duality; or simply by digging deeper and deeper into layers of subquark physics, new successes may be achieved sooner or later, and then as a consequence the harmonious relation with the sponsor, which existed in a long period after the world war II, will be restored. Or, on the other hand, the community can give up the strong version of reductionism, and assume an endless tower of effective theories, in which each theory is a particular response to a particular experimental situation, and none of them can ultimately be regarded as the fundamental theory.

The second direction may not mean to take a purely phenomenological approach, since the community can still endorse Philip Anderson's idea of fundamentality. That is, there are many fundamental theories of physics, each of which is responsible for a certain level of complexity in the physical world, no one can claim to be more fundamental than others. Thus the door is open for understanding much wider range of complex systems which might have more direct relevance to social benefits and human life. Anderson's idea has given the high energy physics community intellectual resources for defining its true in the society, providing new justifications for its activities, and taking its moral responsibilities into account. For some diehard foundationalists, however, the second perspective is unacceptable because it is incompatible with the quasi-religious thrill they are seeking for in their research activities.

III. Now let me say a few words about the future of the standard model. Briefly put, there are three possibilities. If Einstein is right and the rock-bottom final truths of nature are accessible and attainable, then the standard model may be developed into a consistent fundamental theory. That is, all the challenging difficulties we mentioned above will be solved sooner or later without altering its underlying assumptions and basic structures, along the lines either of constructive field theory or of algebraic quantum field theory. Some mathematical physicists, such as Wightman and Jaffe, firmly believe in this possibility. Others find it difficult to conceive such a possibility, mainly because of the necessity in revising our classical conception of the spacetime underpinning of field theory, within which the standard model is constructed. If we want a consistent theory, consistent with both general relativity and quantum theory, then spacetime must be curved and/or quantized, then the foundations of quantum field theory, such as locality, causality, the vacuum, particles, gauge conditions, or even the very idea of spacetime itself, have to be radically revised. After these revisions, even if they are carried out successfully, the resulting theory would hardly be recognizable as a successor of the standard model.

The second possibility, still in accordance with the Einsteinian perspective that final truths of nature are accessible and attainable, is that the standard model will appear as a limiting case, as an effective theory, of this final theory, which itself may not be a field theory anymore, but perhaps is something like a string theory. All string theorists plus some other enthusiasts in string approach, such as Weinberg and Coleman, take this as the only possibility for the future status of the standard model. Many physicists find it hard to swallow because, first, the only visible candidate for the final theory now, the string theory, is hardly a physical theory in its conventional sense; and second, the very idea of having a final theory is philosophically and also psychologically unacceptable, because this would imply an end to theoretical physics, although may not be end to applied physics. The true meaning of this possibility, however, is a rejection of the standard model as a fundamental theory, and its conception is a response to the long period of stagnation after the advent of the standard model, and to the conceptual crisis facing the standard model as we sketched above. The very appearance of string itself is a manifestation of the crisis that the standard model could offer no theoretical resource for dealing with challenging questions, most important among them are gravity, divergence, and unification.

If Einstein's conviction turns out to be wrong, and Philip Anderson's conception of a multi-foundational reality, which is in harmony with the world picture suggested by a radical interpretation of effective field theory, is accepted by the high energy physics community and taken to be a guidance in their theoretical endeavor, then the standard model will appear to be one effective theory in an endless tower of effective theories. Some physicists may argue that this third possibility sounds suggesting an infinite onion and thus is quite boring, and is very unlikely to occur because the necessity of radical revision of the foundations of theoretical physics. But if we remove the misconception that all effective theories must have same foundations and similar structures, and insist only that no theory would be a final theory to end physics, thus all theories, regardless their foundational situation and structural features, must in some sense be an effective theory, then, perhaps this may be the possibility for the future status of the standard model.