The conceptual foundations and the philosophical aspects of renormalization theory

Article *in* Synthese · October 1993

DOI: 10.1007/BF01255832

2 authors:

Tian yu Cao
Boston University

Silvan S. Schweber
Brandeis University

TIAN YU CAO AND SILVAN S. SCHWEBER

THE CONCEPTUAL FOUNDATIONS AND THE PHILOSOPHICAL ASPECTS OF RENORMALIZATION THEORY*

1. INTRODUCTION

As a representation for describing subatomic entities and as a framework for the hierarchical structure that can be built from these entities (mesons, nucleons, nuclei, atoms, etc.), quantum field theory (QFT) embodies a reductionist view of science. Serious doubt has often been cast on the whole program, particularly when the foundations themselves were found to be in a state of confusion. During the 1930s and most of the 1940s the infinite results that occurred within the framework of QFT in higher-order perturbative calculations made most physicists doubt the stability of the foundations of QFT.[1] The doubts were dispelled for a while after a renormalization procedure was proposed and carried out in 1947–48 by Kramers (in Schweber, 1985), Bethe (1947), Lewis (1948), Schwinger (1948a, 1948b), Tomonaga (1946), and Feynman (1948a, 1948b, 1948c), and spectacular successes were achieved in explaining and predicting radiative corrections to electromagnetic processes. More specifically, confidence in QFT got a further boost when a proof of the renormalizability of the *S*-matrix in quantum electrodynamics (QED) – the simplest case of a QFT – was suggested by Dyson (1949a, 1949b; cf. Schweber, 1986a, 1986b). However, Dyson's proof was not conclusive. Some loopholes – such as those related to the overlapping divergences and the lack of a rigorous proof of the convergence of the renormalization procedure in each order – were later closed by Ward (1950, 1951), Salam (1951a, 1951b), Weinberg (1960), and Mills and Yang (1966). Others have persisted, most notably, the nonconvergence of the power series used in the perturbative expansion of the *S*-matrix in terms of which Dyson's and Weinberg's proof are formulated – as pointed out by Dyson (1952), Hurst (1952), and Thirring (1953). Soon after Dyson's proof was published, serious arguments challenging the stability of renormalization theory at various

Synthese **97**: 33–108, 1993.

levels – mathematical, physical, and conceptual – were advanced by Dirac (1969a, 1969b, 1973a, 1973b, 1983), Schwinger (1970, 1973, 1983), Källen (1953), Landau et al. (1954a, 1954b, 1954c, 1954d, 1956), Landau (1955), Landau and Pomeranchuck (1955), and others, and even by Dyson (1952) himself. There were various responses to the conceptual difficulties of renormalization theory. At one end of the spectrum were the axiomatic field theorists, who sought to clarify the theoretical structure of QFT and to construct a stable renormalized theory of quantum fields (cf. Streater and Wightman, 1964; Velo and Wightman, 1973; Wightman, 1986). At the other end were Landau, Chew, and other S-matrix theorists, who denounced the whole framework of QFT, including its renormalized version, not merely because of its empirical failure in describing the strong and weak interactions, but principally because of its conceptual instability (cf. Cushing, 1990; Cao, 1991).

The basic question that had to be answered was: How can one account for the great empirical success of renormalized QED, *a conceptually unstable theory*?[2] A stubborn logician confronting this paradox would probably reject either renormalization theory or field theory (or both!), and some physicists such as Dirac, Landau, and Chew did so. But most field theorists reasoned differently. Ignoring the stability problem in renormalization theory, they argued that if meaningful calculations could only be carried out within the framework of renormalized perturbative theory, then in fact renormalizability should be taken as a crucial constraint on theory construction. It is a historical fact that the further developments of QFT beyond the scope of QED have been accomplished using the principle of renormalizability as a guideline. The most convincing case in point is Weinberg's unified field theory of the electroweak interactions. As Weinberg (1980a) remarked in his Nobel lecture, if he had not been guided by the principle of renormalizability, his theory of electroweak interactions would have received contributions not only from $SU(2) \times U(1)$-invariant vector boson exchanges – which were believed to be renormalizable, though not proven to be so until few years later by 't Hooft (1971a, 1971b) and others (Lee and Zinn-Justin, 1972; 't Hooft and Veltman, 1972a, 1972b, 1972c; Becchi et al., 1974) – but also from $SU(2) \times U(1)$-invariant four-fermion couplings, which were known to be nonrenormalizable, and the theory would have lost most of its predictive power.

Certainly this historical fact deserves philosophical reflection, particularly with reference to the criteria of theory appraisal, theory acceptance, and theory choice.

During the mid 1970s the fundamental nature and essential character of the renormalizability principle began to be challenged. As a result of two decades of fruitful interactions between QFT and statistical mechanics, the understanding by theoretical physicists of certain foundational aspects of renormalization theory underwent a radical transformation. At the heart of the transformation was the emergence of the new concept of 'broken scale invariance' and the related renormalization group approach. Weinberg (1978) was one of the first to assimilate the physical insights developed principally by K. G. Wilson (Wilson and Kogut, 1974; Wilson, 1975) in context of critical phenomena – for example, the existence of the fixed point solutions of renormalization group equations and the conditions for trajectories in coupling-constant space passing through fixed points – and to apply them within the context of QFT. His intention was to explain or even replace the renormalizability principle with a more fundamental guiding principle, which he labeled "asymptotical safety". Yet this program was soon to be overshadowed by another, that of 'effective field theory' (EFT), also initiated by Weinberg (1979, 1980b). At first, EFT was a less ambitious program than that encompassed by asymptotically safe theories, because EFT still takes renormalizability as its conceptual basis. EFT, however, has led to a radical change of outlook, together with a thorough examination of the very concept of renormalizability and a clarification of the ontological basis of QFT (cf. Schweber, 1993).

The present paper is an introduction to a comprehensive account of the history of renormalization theory, which the authors have undertaken and is still in progress. After giving the necessary background (Section 2), we examine the foundations (Section 3) and the philosophical aspects (Section 4) of renormalization theory. With the help of a brief analysis of the nature of physical theories in general, and of the structure of the hypothetico-deductive method adopted by QFT in particular, the bearings of this case study on more general topics in the philosophy of science, such as the criteria for theory acceptance and theory choice, and the issue of 'realism versus instrumentalism', are discussed in some detail in the last section.

2. BACKGROUND

Renormalization theory is a complicated conceptual system. It can be understood from various perspectives. First, the renormalization procedure can be viewed as a technical device for circumventing – i.e., isolating and discarding – the infinite results that occur in QFT perturbative calculations.[3] Second, the concept of renormalization helps to clarify the conceptual basis of QFT and, it is hoped, to establish its stability. Third, renormalizability can also be elevated to the status of a regulative principle, guiding theory construction and theory selection within the general framework of QFT.

Historically, the emergence of renormalization theory in the late 1940s was a response to the divergence difficulties of QFT. In its original formulation, renormalization theory was technical and conservative in character, and it is important to keep these initial characteristics in mind when trying to understand its further developments. It was technical because it involved a series of algorithmic steps for obtaining numerical results from the theory, numbers that could be compared to experimental data, for example, the Lamb shift (Lamb and Retherford, 1947) and the anomalous magnetic moment of the electron (Nafe et al., 1947). It was conservative because it took the framework of QFT as given, and made no attempts to alter its foundations. In fact, Dyson took its conservative character as one of its endearing features (Schweber, 1986a). Thus a brief introduction to the conceptual framework of QFT is in order.

QFT is a system consisting of local field operators that obey equations of motion, certain canonical commutation and anticommutation relations (for bosons and fermions, respectively), and a Hilbert space of state vectors that is obtained by the successive application of the field operators to the vacuum state, which is assumed to be unique. Let us look at three of the assumptions that are involved in greater detail.

(i) *The locality assumption.* In QFT this asserts that field operators on a spacelike surface commute (bosons) or anticommute (fermions) with each other. The assumption is a legacy of the point model of particles and its description of interactions among them. At first glance, locality seems merely to be a statement of the rejection of the possibility of action-at-a-distance, and a means to keep the representation in compliance with special relativity.[4] But an examination of the construction of the point model of the electron reveals that it is also an attempt to resolve a difficulty in Lorentz's theory of the electron (1904a, 1904b). According to J. J. Thomson (1881), the energy contained in the field of a spherical charge of radius a is proportional to $e^2/2a$. Thus, when the radius of a Lorentzian electron goes to zero, the energy diverges linearly. But if the electron is given a finite radius, then the repulsive Coulomb force within the sphere of the electron makes the configuration unstable. Poincaré's response (1906) to the paradox was the suggestion that there might exist a nonelectromagnetic cohesive force inside the electron to balance the Coulomb force, so that the electron would not be unstable. Two elements of the model have exercised great influence on later generations: (a) the notion that the mass of the electron has, at least partly, a nonelectromagnetic origin; and (b) the notion that the nonelectromagnetic compensative interaction, when combined with the electromagnetic interaction, would lead to the observable mass of the electron. Thus, the point of departure of Stueckelberg (1938), Bopp (1940), Pais (1945), Sakata (1947), and many others in their studies of the problem of the electron's self-energy is Poincaré's ideas.

The equilibrium of the Poincaré electron is not stable against deformations. This was first pointed out by Fermi in 1922 (cf. Rohrlich, 1973), and this observation elicited another kind of response to the difficulty, first stated by Frenkel (1925). Frenkel argued that since the electron is elementary and has no substructure, the inner equilibrium of an extended electron is a meaningless problem within the classical framework. By adopting the point model, Frenkel eliminated the 'self-interaction' between the parts of an electron – and thus the stability problem – but he could not eliminate the 'self-interaction' between the point-electron and the electromagnetic field it produces without abandoning Maxwell's theory. The problem Frenkel left open became more acute when QFT came into being.

Frenkel's idea of the point-electron[5] was quickly accepted by physicists and became the conceptual basis for QFT. The idea of looking for a structure of the electron was given up because, as Dirac (1938, p. 950) suggested, "the electron is too simple a thing for the question of the laws governing its structure to arise". It is clear, therefore, that what is hidden in the locality assumption is an acknowledgment of our ignorance of the structure of the electron and that of other elementary entities described by QFT. The justification given for the point model and the consequent locality assumption is that they constitute good, approximate representations at the energies available in present experiments, energies that are too low to explore the inner structure of the particles.

(ii) *The operator field assumption.* When Jordan (in Born et al., 1926) and Dirac (1927a, 1927b) extended the methods of quantum mechanics to electromagnetism, the electromagnetic field components were promoted from classical commuting variables to quantum mechanical operators. The same procedure could also be applied to fields describing fermions (Jordan and Klein, 1927; Jordan and Wigner, 1928; cf. Darrigol, 1986). These local field operators have a direct physical interpretation in terms of the emission and absorption and the creation and annihilation of the quanta associated with the particles. The creation of a particle is realized as a localized excitation of the vacuum. According to the uncertainty principle, a localized excitation implies that arbitrary amounts of energy and momentum are available for the creation of particles. Thus the result of applying a field operator to the vacuum state is not a state containing a single particle, but rather results in a superposition of states containing arbitrary numbers of particles that is constrained only by the conservation of the relevant quantum numbers.[6] An operator field is defined by the totality of its matrix elements, hence, it should be clear that an overwhelming proportion of these refer to energies and momenta that are far outside experimental experience.

(iii) *The plenum assumption of the bare vacuum.* The ontological status of the 'bare' vacuum has been a widely discussed subject. It may be recalled that Furry and Oppenheimer (1934), Pauli and Weisskopf (1934), Wentzel (1943), and others raised objections to Dirac's idea (1930) of the vacuum being a state in which all the (one-particle) negative energy states were filled. The strongest argument put forth against the plenum assumption was the following: according to special relativity, the vacuum must be a Lorentz-invariant state of zero energy, zero momentum, zero angular momentum, zero charge, zero whatever, that is, a state of nothingness (cf. Weisskopf, 1983, p. 69). However, when certain phenomena supposed to be caused by the vacuum fluctuations were analyzed, the very same physicists who objected to the plenum assumption tacitly took the vacuum as something substantial, namely as a polarizable medium, or assumed it to be an underlying substratum, the scene of wild activities. In other words, they actually adopted the plenum assumption.

The local coupling among quanta and the fact that the application of field operators on the vacuum results in strictly local excitations imply that in QFT calculations one has to consider virtual processes involving arbitrarily high energy. However, except for the consequences imposed by such general constraints as unitarity, there exists essentially no empirical evidence for believing the correctness of the theory at these energies. Mathematically, the inclusion of these virtual processes at arbitrarily high energy results in infinite quantities that are obviously undefinable. Thus the divergence difficulties are not external. They are internal to the very nature of QFT: they are constitutive within the canonical formulation of QFT. In this sense the occurrence of the divergences clearly pointed to a deep instability in the conceptual structure of QFT.

In addition to various proposals for radically altering the foundations of QFT, two different responses were advanced to overcome this instability. The first one was developed independently by Pais and by Sakata, and was in the spirit of Poincaré's solution to the stability problem of the Lorentz electron. It put forth the idea of compensation: fields of unknown particles were introduced in such a way as to cancel the divergences produced by the known interactions. The second response was the renormalization program, which is the central subject of the present paper.

In the 1930s, Dirac (1934), Heisenberg (1934), Weisskopf (1936), Kramers (1938), and others had already put forth the idea of renormalization in terms of subtractions. But it required the precise experimental findings on the spectrum of hydrogen and deuterium that were obtained using techniques and instruments developed during World War II to stimulate the further elaboration of this idea. The explanation of the accurate and reliable data obtained by Lamb and Retherford in their measurements of the fine structure of hydrogen and of Rabi's results on the hyperfine structure of hydrogen and deuterium became an outstanding challenge for theoretical physicists. In the process they developed algorithms for obtaining finite numbers for the measured quantities in their QFT-based calculations, and put forth suggestive ideas for justifying the algorithms (Schweber, 1986a).

The ideas and algorithms developed by Kramers, Bethe, Lewis, Schwinger, and Tomonaga can be summarized as follows:

(i) The divergent terms that occur in the QED calculations are identifiable in a Lorentz- and gauge-invariant manner, and can be interpreted as modifying the mass and charge parameters that are introduced in the original Lagrangian.

(ii) By identifying the modified, or renormalized, mass and charge parameters with the physically observable masses and charges of physical particles, all the divergences are absorbed into the mass and charge renormalization factors, and finite results in good agreement with experiments are obtained.[7] Thus the measurements of Lamb and of Rabi can be explained within the framework of (renormalized) QED.

The crucial assumption underlying the whole renormalization program was first expressed succinctly by Lewis (1948, p. 173):

> The electromagnetic mass of the electron is a small effect and . . . its apparent divergence arises from a failure of present day quantum electrodynamics above certain frequencies. . . ,

and somewhat more fully by Schwinger (1948a, p. 416) in his first paper on QED:

> Electrodynamics unquestionably requires revision at ultra-high energies, but is presumably accurate at moderate relativistic energies. It would be desirable, therefore, to isolate those aspects of the current theory that essentially involve high energies, and are subject to modification by a more satisfactory theory, from aspects that involve only moderate energy and are thus relatively trustworthy.

It is clear that only when a physical parameter (which when calculated in perturbation in QFT may turn out to be divergent) is actually finite and small can its separation and amalgamation into the 'bare' parameters be regarded as mathematically justifiable. The failure of QFT at ultra-relativistic energies, as indicated by the divergences in perturbation theory, implied that the region in which the existing framework of QFT is valid should be separated from the region in which it is not valid and in which new physics would become manifest. It is impossible to determine where the boundary is, and one does not know what theory can be used to calculate the small effects that are not calculable in QFT. However, this separation of knowable from unknowable, which is realized mathematically by the introduction of a cutoff, can be schematized by using the phenomenological parameters that must include these small effects.

Neither Lewis nor Schwinger nor Tomonaga made explicit use of a cutoff. They directly identified the divergent terms with corrections to the mass and the charge, and removed them from the expressions for real processes by redefining the masses and charges. By contrast, Feynman's efficient calculational algorithm (1948b, 1948c; cf. Schweber, 1986b) is based on the explicit use of a relativistic cutoff. The latter consists of a set of rules for regularization, which makes it possible to calculate physical quantities in a relativistically and gauge-invariant manner, but still results in divergent expressions in the limit as the cutoff mass goes to infinity. With a finite cutoff, this artifice transforms essentially purely formal manipulations of divergent quantities, i.e., the redefinition of parameters, into quasi respectable mathematical operations. If, after the redefinition of mass and charge, other processes are insensitive to the value of the cutoff, then a renormalized theory can be defined by letting the cutoff go to infinity. A theory is called renormalizable if a finite number of parameters are sufficient to define it as a renormalized one.

Physically, Feynman's relativistic cutoff is equivalent to introducing an auxiliary field (and its associated particle) to cancel the infinite contributions due to the ('real') particles of the original field. Feynman's approach is different from *realistic* theories of regularization or compensation. In the latter, auxiliary particles with finite masses and positive energies are assumed to be observable in principle, and are described by field operators that enter the Hamiltonian explicitly. Feynman's theory of a cutoff is *formalistic* in the sense that: (i) the auxiliary masses are used merely as mathematical parameters, which finally tend to infinity and are nonobservable in principle; and (ii) the coupling constant associated with the auxiliary particle would be imaginary.[8] Representative of the 'realistic' approach are the papers of Sakata (1947, 1950; see also Sakata and Hara, 1947; Sakata and Umezawa, 1950), Umezawa (Umezawa et al., 1948; Umezawa and Kawabe, 1949a, 1949b), and other Japanese physicists, as well as that of Rayski (1948). Among the 'formalists' we find, in addition to Feynman, Rivier and Stueckelberg (1948) and Pauli and Villars (1949).

It was Dyson who, as a synthesizer, showed that Feynman's results and insights were derivable from Tomonaga's and Schwinger's formulation of QED. Furthermore, Dyson was able to outline a proof of the renormalizability of QED. Renormalizability meant that mass and charge renormalization removed all the divergences from the S-matrix of QED to all orders of perturbation theory. He also suggested that renormalized QED might well be a quasi-stable theory. Empirically, the renormalized version of QED has enjoyed great success because of its astonishing predictive power, both in calculating the anomalous magnetic moment of the electron and the Lamb shift in hydrogen, and in estimating the radiative corrections to high energy electron-electron and electron-positron scattering. In the late 1940s and early 1950s it was hoped that, by successfully circumventing the divergence difficulties, a quasi-stable framework of QFT could be constructed, and moreover, as Pauli suggested, that it might fix the masses and charges of the particles that appear in the theory. This turned out to be too optimistic, although some advances toward a rigorous proof of the renormalizability of QED were made. Crucial among these were:

(i) the solutions to the overlapping divergences given by Salam, Ward, and Mills and Yang; and
(ii) the convergence theorem of Weinberg, which is necessary for the proof that in renormalizable theories all the ultraviolet divergences do cancel to all orders of perturbation theory, despite the occurrence of complicated divergent subgraphs.

However, even though renormalizability had been accepted as a property of QED, why the renormalization program actually works in QED remained conceptually quite unclear. This question can be divided into two parts:

(i) Why do the apparent divergences arising from a failure of unrenormalized QED above certain energies actually give rise to small effects?
(ii) Why are the representations of nature by renormalized theories stable, and more specifically why are they so very insensitive to whatever happens at very high energy?

While some progress toward an answer to the second part of the question has been made during the past two decades,[9] no real insight has been obtained toward being able to give an answer to question (i) during the more than four decades since Lewis first stated the smallness assumption. But perhaps an even more fundamental question regarding renormalization theory was whether all the interactions in nature are renormalizable.

Dyson was aware that the answer to the question was negative, and so reported to the Olstone conference (Schweber, 1986a). Detailed, explicit, negative examples were given immediately after Dyson's classic papers on renormalization appeared. For example, Feldman (1949) observed that the electromagnetic interactions of the vector meson are nonrenormalizable. Kamefuchi (1951) pointed out that the Fermi four-fermion direct interactions are also nonrenormalizable, and Peterman and Stueckelberg (1951) noted that the interaction of a magnetic moment with the electromagnetic field (a Pauli term of the form $f\psi\sigma^{\mu\nu}\psi F_{\mu\nu}$) is likewise nonrenormalizable. Thus the unavoidable question arose: Should nature be described only by renormalizable theories? For physicists, such as Bethe, who had elevated renormalizability from a property of QED to a regulative principle guiding theory selection, the answer was affirmative (see Schweber et al., 1955). They justified their position in terms of predictive power. They argued that since the aim of fundamental physics is to formulate theories that possess considerable predictive power, "fundamental laws" must contain only a *finite* number of parameters. Only renormalizable theories are consistent with this requirement. While the divergences of nonrenormalizable theories could possibly be eliminated by absorbing them into appropriately specified parameters, an *infinite* number of parameters would be required and such theories would initially be defined with an infinite number of parameters appearing in the Lagrangian.

According to the renormalizability principle, the interaction Lagrangian of a charged spin 1/2 particle interacting with the electromagnetic field cannot contain a Pauli moment. Similarly, a pseudovector coupling of the pion to the nucleon was excluded. By the same reasoning, Fermi's theory of weak interaction lost its status as a fundamental theory. A more complicated application of the renormalization constraint was the rejection of the pseudoscalar coupling of pions to nucleons in the strong interactions. Formally, the pseudoscalar coupling was renormalizable. Yet its renormalizability was not realizable because the radiative corrections it produces are too large to justify the use of perturbation theory – which is the only framework within which the renormalization procedure works. This contributed to the popularity of the dispersion relations approach and to the adoption of Chew's S-matrix theory approach, which rejected the whole framework of QFT, by a considerable number of theorists.[10]

The case of the Yang–Mills field (1954a, 1954b) deserves special attention. Physicists were interested in the original Yang–Mills theory in part because it was conjectured to be renormalizable, even though the massless bosons required by the gauge invariance could not be responsible for short-range nuclear forces. The massive version of Yang–Mills theory was unacceptable because the massive gauge bosons spoiled not only the gauge invariance but the renormalizability of the theory as well. Gell-Mann was attracted by the theory and tried to find a "soft-mass" mechanism that would allow the renormalizability of the massless theory to persist in the presence of gauge boson masses, but he did not succeed (cf. Gell-Mann, 1987, 1989). Where Gell-Mann failed, Weinberg (1967), Salam (1968), and Gross and Wilczek (1973a, 1973b) succeeded with the help of the Higgs mechanism and of renormalization group equations, respectively. After the apparent proof of the renormalizability of a Yang–Mills theory,[11] the latter theory became the paradigmatic case of QFT and constituted an extension of Dyson's original program into a new area.

It would not be too great an exaggeration to claim that the most substantial advances in QFT that have been achieved in the past four decades have been guided and constrained by the renormalizability principle. It is certainly true that renormalization theory saved QFT, made it manipulable, and allowed one to calculate with it, and thus revived the faith of theorists in QFT (Schweber, 1986a). Be that as it may, no consensus has ever been reached as to whether renormalizability is an essential characteristic of QFT or a universal principle constraining all the possible descriptions of nature. As a matter of fact, since the early 1950s serious arguments have been advanced challenging the stability of renormalization theory and casting doubts on the foundations of QFT. The debate has led to a deeper understanding of the physics and the philosophy of renormalization, and has helped to clarify the foundations of QFT. This can be viewed as another way in which renormalization theory has advanced QFT into a new phase.

An unrenormalized theory is certainly unstable due to the presence of ultraviolet divergences and by virtue of the infinities that stem from the infinite volume of spacetime. The latter have to be disentangled from the former and excludes the Fock representation as a candidate for the Weyl form of the canonical commutation relations. The occurrence of these two kinds of infinities makes it impossible to define a Hamiltonian operator, and the whole scheme of canonical quantization of QFT collapses.

The stability problem in renormalized theories is very different from that of unrenormalized ones. The ultraviolet divergences are supposed to be circumventable by the renormalization procedure. Some of the remaining difficulties, such as how to define local fields and their equivalence class, and how to specify asymptotic conditions and the associated reduction formula, were analyzed in a rigorous fashion by axiomatic field theorists with the help of distribution theory and normed algebra (cf. Wightman, 1989). Yet new problems created by renormalization theory invited serious criticisms that were advanced on three different levels.

At the mathematical level, Dirac (1969b) criticized renormalization theory for neglecting infinities instead of infinitesimals, a procedure radically at odds with the usual custom in mathematics. Lewis's smallness assumption anticipated and seems to invalidate Dirac's criticism, but the assumption itself has to be justified in the first place.[12]

At the physical level, Heitler (1961) and others noted that the mass differences of particles (such as the pions and the nucleons), which are identical except for their electric charge, could not be calculated using renormalization theory. It is not difficult to establish that if the mass differences are of electromagnetic origin, then the divergent electromagnetic self-energy will lead to infinite mass differences. This difficulty clearly indicated that renormalization theory could not fulfill Pauli's hope that it would provide a general theory to account for the mass ratios of the "elementary particles". In addition, renormalization theory was criticized as being too narrow a framework to accommodate the representations of such important phenomena as the CP-violating weak interactions and the gravitational interactions. But the gravest defect of renormalization theory was made manifest around 1970, when it was recognized that it is in direct and irreconcilable conflict with the chiral and the trace anomalies that occur in high orders of QFT, ironically, as a consequence of the demand of renormalization (cf. Jackiw, 1972).

At the conceptual level, the stability of renormalization theory was challenged by Dyson, Källen, Landau, and others. In 1953, Källen claimed to be able to show that, starting with the assumption that all renormalization constants are finite, at least one of the renormalization constants in QED must be infinite. For several years this contradictory result was accepted by most physicists as evidence for the inconsistency of QED. However, as was later pointed out by some critics (e.g., Gasiorowicz et al., 1959), his results depended on some notoriously treacherous arguments involving interchanges of the orders of integration and summation over an infinite number of states, and was thus inconclusive. Källen himself later acknowledged this ambiguity (1966).

More serious arguments challenging the stability of renormalization theory were expressed in terms of the breakdown of perturbation theory. As is well known, Dyson's renormalization theory was only formulated within the framework of perturbation theory. The output of perturbative renormalization theory is a set of well-defined formal power series for the Green functions of a field theory. However, it was soon realized that these series – and in particular the one for the S-matrix – were most likely divergent. Thus theorists were thrown into a state of confusion and could not give an answer to the question: In what sense does the perturbative series of a field theory define a solution? Interestingly enough, the first theorist to be disillusioned by perturbative renormalization theory was Dyson himself. In 1952, Dyson gave an ingenious argument that suggested that after renormalization all the power series expansions were divergent. The subsequent discussion by Hurst (1952), Thirring (1953), Peterman (1953a, 1953b), Jaffe (1965), and other axiomatic and constructive field theorists added further weight to the assertion that the perturbative series of most renormalized field theories diverge, even though there is still no complete proof in most cases.

A divergent perturbative series for a Green function may still be asymptotic to a solution of the theory. In the mid 1970s the existence of solutions for some field theoretical models was established by constructive field theorists and these indicated a posteriori that the solution is uniquely determined by its perturbative expansion (cf. Wightman, 1976). Yet these solutions were exhibited only for field-theoretic models in spacetime continua of two or three dimensions. As for the more realistic four-dimensional QED, in 1952 Hurst had already suggested that the excellent agreement of QED with experiments may indicate that the pertubative series is an asymptotic expansion. However, the investigations of the high energy behavior of QED by Källen, Landau, and especially by Gell-Mann and Low (1954), showed that the perturbative approach in QED unavoidably breaks down, ironically, as a consequence of the necessity of charge renormalization. Landau and his collaborators argued further that remaining within the perturbative framework would lead either to no interaction (zero renormalized charge)[13] or to the occurrence of ghost states rendering the theory apparently inconsistent (Landau, 1955). Both results demonstrated the inapplicability of perturbative theory in renormalized QED.

After the discovery of asymptotic freedom in a wide class of non-Abelian gauge theories, especially in quantum chromodynamics (QCD), the hope was expressed that perturbative QCD would get rid of the Landau ghost and would thus eliminate most doubts as to the consistency of QFT. However, this expectation did not last long. It was soon realized that the ghost that disappeared at high energy reappeared at low energy (cf. Collins, 1984). Thus field theorists were reminded – forcefully and persistently – of the limits of the applicability of perturbative theory. As a result, the stability problem of QFT in general, and the consistency problem of perturbative renormalization theory in particular, was in a state of uncertainty.

The attitude of theoretical physicists toward this issue differed sharply. For most practicing physicists, stability is just a pedantic problem. As pragmatists, they are only guided by their scientific experiences and have little interest in speculating about the ultimate stability of a theory. For Dirac (1963, 1969a, 1969b, 1973a, 1973b, 1983), however, the existing renormalization theory with the cutoff going to infinity was physically illogical and nonsensical. In his opinion, what was required were new forms of interaction and new mathematics, such as the possible use of an indefinite metric (1942), or of nonassociative algebra (1973a), or perhaps something even more esoteric. The positions adopted by Landau and by Chew were more radical and drastic (cf. Cao, 1991). What they rejected were not merely particular forms of interactions and perturbative versions of QFT, but also the paradigm set by QFT, the style of reasoning exemplified by QFT, and the general framework established by QFT. For them the very concept of a local field operator and the postulation of any detailed mechanism for interactions in a microscopic spacetime region were totally unacceptable because these were too speculative to be observable, even in principle. Their position was supported by the presence of divergences in QFT and by the instability or even the inconsistency of the perturbative renormalization theory,[14] even though Landau's zero charge argument could not claim to be conclusive.

The most positive attitude was taken by the axiomatic field theorists, who later called themselves constructive field theorists.[15] In the spirit of Hilbert's tradition they tried to settle the question of the stability of QFT by axiomatization, and took this as the only way to give clear answers to conceptual problems.

While Hilbert tried to legitimize the use of mathematical entities with a proof of the consistency of a formal system consisting of these entities,[16] the axiomatic field theorists went the other way round. They tried to prove the internal consistency of QFT by constructing nontrivial examples whose existence is a consequence of the axioms alone. Without radically altering the foundations of QFT, they tried to overcome the apparent difficulties with its consistency, step by step. Although many of important problems remain open, nowhere did they find any indication that QFT contained basic inconsistencies.

The axiomatic field theorists took the fields to be operator-valued distributions defined with infinitely differentiable test functions of fast decrease at infinity or with test functions having compact support. Essentially, this was a mathematical expression of the physical idea of modifying the exact point model. However, a thus defined theory may still be nonrenormalizable in the sense of perturbation theory. That is, it may still be an unstable theory even though there is no inconsistency involved in it.

Since the mid 1970s, there have been major efforts using the approach of constructive field theory to understand the structure of nonrenormalizable theories and to establish the conditions under which a nonrenormalizable theory can make sense. One of the striking results of this enterprise is that the solutions of some nonrenormalizable theories have only a finite number of arbitrary parameters. This is contrary to their description in terms of the perturbative series. It has been speculated that the necessity for an infinite number of parameters to be renormalized in perturbation theory may come from an illegitimate power series expansion (cf. Wightman, 1986). It is certainly the case that in these efforts the axiomatic and the constructive field theorists have exhibited openness and a flexible frame of mind. Yet future developments in understanding the foundations and proving the consistency and stability of renormalization theory may involve changes in some assumptions that have not yet been challenged, and that have not been captured by any axiomatization of the present theory. In any case, the failure to construct a soluble four-dimensional field theory, despite intensive efforts for nearly four decades, indicates that the axiomatic and the constructive field theorists are meeting considerable difficulty in solving the consistency problem of QFT in Hilbert's sense, let alone its stability. It also has dampened their initial optimism somewhat.

The finitists, among whom are Salam (1973; Salam and Strathdee, 1970) and the various advocates of supergravity (e.g., Hawking, 1980) and superstrings (e.g., Green et al., 1987), are more optimistic than the axiomatic and the constructive field theorists. Their hope is that by including gravitational interactions in the existing formulations of QFT systems, it will be possible to construct a finite theory without any infinite renormalizations. They thus hope to solve the stability problem of QFT without involving any renormalization. It should, however, be recalled that hopes come and go, and moreover that they seem to be short-lived.

Two additional contrary views on renormalization, advanced respectively by Sakata (1950, 1956; Sakata and Umezawa, 1950; Sakata et al., 1952) and by Schwinger (1970, 1973, 1983), were expressed in terms of their concerns regarding the structure of the "elementary" particles. For Sakata, renormalization theory was only an abstract formalism, behind which lay hidden the concrete structure of elementary particles. His position was that when renormalization theory would encounter a defect (and its limitations would be exposed), it would become necessary to look for and to analyze more closely the structure of the elementary particles. Under Sakata's influence, more efforts were invested in Japan in model-building of the constituents of elementary particles than in the analysis of the theoretical structure of QFT. As a result, little emphasis was placed on renormalization theory as an essential conceptual ingredient of QFT (cf. Takabayasi, 1983; Aramaki, 1989).

Schwinger's views of renormalization are of particular interest, not merely because he is one of the founders of renormalization theory, but principally because he has given penetrating analyses of the philosophy of the renormalization program and is one of its most incisive critics. According to Schwinger, the unrenormalized description, which adopts local field operators as its conceptual basis, contains speculative assumptions about the dynamic structure of the physical particles that are sensitive to details at high energy. However, we have no reason to believe that the theory is correct in that domain. In accordance with Kramers's precept that QFT should have a structure-independent character, which Schwinger accepted as a guiding principle, the renormalization procedure that he elaborated removed any reference to very high energy processes and the related small distance and inner structure assumptions. He thus shifted the focus from the hypothetical world of localized excitations and interactions to the observed world of physical particles. But Schwinger found it unacceptable to proceed in this tortuous manner of first introducing physically extraneous structural assumptions, only to delete them at the end in order to obtain physically meaningful results. This constitutes a rejection of the philosophy of renormalization. But renormalization is essential and unavoidable in a local operator field theory if the latter is to make any sense. In order to bring his criticism to its logical conclusion, Schwinger introduced numerically valued (nonoperator) sources and numerical fields to replace the local field operators. These sources symbolize the interventions that constitute measurements of the physical system. Furthermore, all the matrix elements of the associated fields, the operator field equations, and the commutation relations can be expressed in terms of the sources. An action principle gives succinct expression to the formalism. According to Schwinger, his source theory takes finite quantities as primary, and it is thus free of divergences. This theory is also sufficiently malleable to be able to incorporate new experimental results, and to extrapolate them in a reasonable manner. Most important, it can do so without falling into the trap of having to extend the theory to arbitrarily high energies – which constitute unexplored domains where new, unknown physics is sure to be encountered.

Thus in Schwinger's approach the ultimate fate of renormalization theory is for it to be eliminated and excluded from any description of nature. He tries to implement this by abandoning the concept of a local operator field, which constitutes a drastic alteration of the foundations of QFT. The radical character of Schwinger's approach, whose foundations were laid in his 1951 paper and elaborated in the 1960s and 1970s, was not recognized until the mid 1970s when the renormalizability principle was first challenged. By that time new, important insights into renormalization and renormalizability had been gleaned from studies using renormalization group methods, resulting in a new understanding of renormalization and of QFT, and also in novel attitudes toward scientific theories in general. Thus a renewed interest in nonrenormalizable theories manifested, and the 'effective field theory' approach began to gain its popularity. In this changed conceptual context, the most perspicacious theorists – e.g., Weinberg (1979) – began to realize that Schwinger's ideas were essential in the radical shift of outlook in fundamental physics.

3. TRANSFORMATIONS OF FOUNDATIONS

The conceptual foundations of renormalization theory have undergone a radical transformation during the last four decades. These changes are the result both of attempts to solve conceptual anomalies within the theory itself and of fruitful interactions between QFT and statistical physics. Implicit assumptions concerning such concepts as regularization, cutoff, dimensionality, symmetry, and renormalizability have been clarified, and the original understanding of these concepts is being transformed. New concepts of symmetry-breaking, either spontaneous or anomalous, of renormalization group transformations, of decoupling of high energy processes from low energy phenomena, of sensible nonrenormalizable theories, and of effective field theories have been developed, drawing heavily on dramatic progress in statistical physics. As a result of these advances there has emerged a new understanding of renormalization, a clarification of the theoretical structure of QFT and its ontological basis, and, most importantly, a crucial shift of outlook in fundamental physics. Section 3.1 examines these foundational transformations, whose philosophical implications will be discussed in 3.2.

3.1. *Cutoff*

As noted in the last section, the renormalization procedure consists essentially of two steps:

(i) for a given theory (e.g., QED or the Weinberg–Salam theory of the electroweak interactions), an algorithm is specified for an unambiguous separation of the ascertainable low energy processes from the high energy processes that are not known, the latter being describable only by new future theories; and

(ii) the incorporation of the effect of the neglected high energy processes on the physics that is described by the theory is accomplished by a redefinition of the finite number of parameters of the theory.

The redefined parameters are not calculable by the theory but can be determined by experiments (Schwinger, 1948a, 1948b). The implicit requirements for the incorporation and redefinition to be possible will be examined in Section 3.2. Here our focus is on the separation.

For Schwinger and Tomonaga, who directly separated the infinite terms by contact transformations, the 'not known' contributions were simply represented by divergent terms that have proper gauge and Lorentz transformation properties and that were properly identified by first constructing an appropriate description of the vacuum state and of the physical one-particle states for the given theory. There were, however, no clues whatsoever in their formulations as to where the boundary separating the knowable from the unknowable energy region lies. It is buried and hidden somewhere in the divergent integrals. Thus the incorporation and redefinition can only be viewed as a species of essentially formalistic manipulations of divergent quantities, with an extremely tenuous logical justification (see Dirac's criticism, 1969a, 1969b).

Feynman, Pauli and Villars, and most other physicists, took an approach that differed from Schwinger's and Tomonaga's. They temporarily modified the theory with the help of a regularization procedure so to make the integrals finite. In the momentum cutoff regularization scheme introduced by Feynman (1948c) and by Pauli and Villars (1949), the boundary line separating the knowable region from the unknowable is clearly indicated by the momentum cutoff introduced.[17] Below the cutoff, the theory is supposed to be trustworthy, and the integrals for the higher-order corrections can be justifiably manipulated and calculated. The unknown high energy processes that occur above the cutoff are excluded from consideration as they have to be. Up to this point, Feynman's scheme seems superior to Schwinger's in implementing the basic ideas of renormalization, which were first clearly stated by Schwinger. It also seems to be more respectable logically and mathematically.

However, the following difficult question must be answered by the various regularization schemes: How are the effects of the excluded high energy processes on the low energy phenomena taken into account? This question is specific to local field theories and is unavoidable within that framework. Feynman's solution, which became the ruling orthodoxy, is to take the cutoff to infinity at the end of the calculation. In this way, all the high energy processes are taken into consideration, and their effects on the low energy phenomena can be incorporated by redefining the parameters that appear in the specification of the theory's

Lagrangian in the manner of Schwinger. The price for accomplishing this is that we can no longer take the cutoff as the threshold energy at which the theory stops being valid and where new theories are required for the correct physical description. Otherwise, we would face a serious conceptual anomaly: taking the cutoff to infinity would mean that the theory is trustworthy everywhere and high energy processes are not unknowable. This is in contradiction with the basic idea of renormalization, and the divergent integrals that result when taking the cutoff to infinity are clear indications that this is not the case.[18]

The implications of taking the cutoff to infinity are very significant. First of all, the boundary line separating the ascertainable and verifiable domain from the unknowable region becomes buried and hidden. It also changes the status of the cutoff from a tentative, and tantalizing, threshold energy to a purely formalistic device, and thus essentially reduces the whole Feynman–Pauli–Villars scheme to Schwinger's original formalistic one. Physically, Feynman's momentum cutoff regularization can be regarded as another, more efficient, formalistic algorithm for manipulating the divergent quantities, which replaces Schwinger's canonical transformations. Or, equivalently, Schwinger's direct identification of the divergent integrals can be viewed as combining Feynman's two steps of introducing a finite cutoff, followed by taking it to infinity. More significantly, taking the cutoff to infinity also reinforces a prevailing formalistic claim that the 'physics' should be cutoff-independent and all explicit reference to the cutoff should be removed on redefining the parameters. The claim seems compelling because the step deprives the cutoff-dependent quantities of any physical meaning. Conversely, the claim in turn allows one to take a purely formalistic interpretation of the cutoff, and forces its removal from real physics.

But what if the cutoff is taken seriously and interpreted realistically as the threshold energy for new physics? Then the orthodox formalistic scheme collapses and the entire perspective changes: the cutoff cannot be taken to infinity and the obverse side of this same coin is that the physics cannot be claimed to be cutoff-independent. In fact, the important advances since the mid 1970s in understanding the physics and the philosophy of renormalization have come from such a realist interpretation (see Polchinski, 1984; Lepage, 1989). There were several intertwined strands of physical reasoning that led to this foundational change, which, in turn, was reinforced by philosophical and practical considerations. The rest of the present section is devoted to disentangling these strands of physical reasoning, leaving the other considerations to be examined in the next section.

To begin with, let us examine the reason why it is possible to take a realist position on the cutoff. As we noted above, the motivation for taking the cutoff to infinity is to take into account the effects of the high energy processes – which are excluded by introducing a finite cutoff – on low energy phenomena. If we can find other ways of retaining these effects while keeping the cutoff finite, then there is no compelling reason for taking the cutoff to infinity. In fact, the realist position has gained adherents since the late 1970s precisely because theorists have gradually come to realize, using detailed power-counting arguments and careful dimensional analysis, that the high energy effects can be retained without taking the cutoff to infinity. This objective can be achieved by adding a finite number of new, local, nonrenormalizable interactions that have the same symmetries as the original Lagrangian, combined with a redefinition of the parameters of the theory (see Wilson, 1983; Symanzik, 1983; Polchinski, 1984; and esp. Lepage, 1989). It is to be noted that the introduction of nonrenormalizable interactions causes no difficulty because the theory has a finite cutoff.

There is a price to be paid for taking this realist position. First, the formalism becomes more complicated by adding the new compensating interactions. The cost is not very high since there are only a finite number of new interactions that need to be added, as these are subject to various constraints. Moreover, this position is conceptually simpler than the formalistic one. Second, the realist formalism is valid only up to the cutoff energy. However, since any experiment can only probe a limited range of energies, this limitation of the realist formalism has actually not caused any real loss in accuracy. Thus the apparent cost is illusionary.

The next question is how to articulate the physical realization of the cutoff so that ways can be found to determine its energy scale. The cutoff in realist theory is no longer a formalistic device or an arbitrary parameter, but acquires physical significance as the embodiment of the hierarchical structure of QFT, and as a boundary separating energy regions that are separately describable by different sets of parameters and different physical laws (interactions) with different symmetries. The discovery of the mechanism of spontaneous symmetry breaking and of the decoupling theorems, to be discussed below, suggests that the value of the cutoff is connected with the masses of heavy bosons, which are associated with the spontaneous symmetry breaking. Since the symmetry breaking makes the otherwise negligible nonrenormalizable interactions[19] detectable due to the absence of all other interactions that are forbidden by the symmetry, the energy scale of the cutoff can be established by measuring the strength of the nonrenormalizable interactions in a theory.

The above discussion has shown in a preliminary fashion that a realist conception of the cutoff is not an untenable position. However, a convincing proof of its viability is possible only when this conception is integrated into a new conceptual network that provides new foundation for understanding renormalizability, nonrenormalizable interactions, and QFT in general. Let us turn to other strands in this network.

3.2. *Symmetry and Symmetry Breaking*

The essential motivation for having a renormalization procedure comes from the necessity of dealing with the divergences that occur in the perturbative solutions of a quantum field theory. In the traditional (formalistic) procedure, after separating the invalid (divergent) parts from the valid (finite) parts of the solutions, the effects of the inaccessible and unknown high energy processes on accessible and knowable low energy phenomena are absorbed by modifying the parameters that enter in the definition of the theory in terms of its Lagrangian. For this amalgamation to be possible, however, the structure of the amplitudes that simulate the unknown and inaccessible high energy dynamics has to be the same as the structure of the amplitudes responsible for the low energy processes. Otherwise, the multiplicative renormalizations would be impossible. To guarantee the required structural similarity, a crucial assumption about the unknown high energy dynamics has to be made that is implicitly built into the very scheme of multiplicative renormalization. This is the assumption that the high energy dynamics is constrained by the same symmetries as those that constrain the low energy dynamics. Now, the solutions of a theory constitute a representation of the symmetry group of the transformations under which the theory is invariant. Therefore, if different symmetries were displayed by the dynamics in different energy regions, this would imply different group-theoretical constraints and a different structure for the solutions in the differing pieces of the dynamics. If this were the case, then the renormalizability of the theory would definitely be spoiled.

In the case of QED, one of the simplest cases, the renormalizability is guaranteed by the somewhat mysterious universality of the $U(1)$ gauge symmetry. However, with the discovery of symmetry breaking, the situation became more complicated. First, in the early 1960s the mechanism of spontaneous symmetry breaking (SSB) was introduced and studied, and then in the late 1960s the phenomenon of anomalous symmetry breaking (ASB) was encountered.[20] These required that the above general consideration about the relationship between symmetry and renormalizability be refined and be made more sophisticated.

Consider SSB. The phenomenon of SSB was first noticed at the beginning of the century (Brown and Cao, 1991), and was rediscovered in the 1950s in investigations of superconductivity. It was explained within the field-theoretical context and integrated into the theoretical structure of QFT by Heisenberg, Nambu, Goldstone, Anderson, Higgs, and others in the early 1960s.[21] In condensed matter and statistical physics, SSB is a statement concerning the properties of the solutions of a dynamical system, namely, that some asymmetrical configurations are energetically more stable than symmetrical ones. Essentially, SSB is concerned with the low energy behavior of the solutions and asserts that some low energy solutions exhibit less symmetry than the symmetry exhibited by the Lagrangian of the system, while others possess the full symmetry of the system. Traced to its foundation, SSB is an inherent property of the dynamical system because the existence and the determination of the asymmetrical solutions are completely determined by the dynamics and the parameters of the system. They are connected to the hierarchical structure of the solution, which, in statistical physics, is manifested in the phenomena of continuous (second-order) phase transitions.

In QFT, SSB makes physical sense only in gauge theories when continuous symmetries are involved. Otherwise, one of its mathematical prediction – namely, the existence of massless Goldstone bosons – would contradict physical observations. Within the framework of gauge theories, all the statements concerning SSB listed in the previous paragraph are valid. There is, in addition to these, another very important assertion that is of relevance to our discussion. In a gauge theory, as for example in the case of the electroweak theory, in contradistinction to the case of explicit symmetry breaking, diverse low energy phenomena can be accommodated in a hierarchy with the help of SSB, without spoiling the renormalizability of the theory. The reason for this is that

SSB affects the structure of physics only at energies lower than the scale at which the symmetry is broken, and thus does not affect the renormalizability of a theory, which is essentially a statement of the high energy behavior of the theory. The profound understanding of these implications of SSB has provided the strong impetus to search for an ultimate unified description of nature, in which natural laws with different invariance properties, symmetrical theories, and asymmetrical physical states all emerge from the highest symmetry that characterizes physics under the conditions present in the early universe, passing through a sequence of phase transitions as the temperature decreases while the universe expands until it reaches the state described by QCD and the electroweak theory.

Such an enterprise has to meet several stringent constraints. One of them arises due to the occurrence of ASB. Generally speaking, ASB is the breakdown of a classical symmetry caused by quantum mechanical effects. It is possible that some symmetries the system possessed in its classical formulation may disappear in its quantified version, because the latter may introduce some symmetry-violating processes. In QFT these arise because of loop corrections, and it is related to the renormalization procedure and the absence of an invariant regulator.

ASB plays an important role in QFT. In particular, the desire to safeguard a symmetry from being anomalously broken can place a very strong constraint on model building. If the symmetries concerned are local, such as gauge symmetries and general covariance, then the occurrence of ASB, which is unavoidable in chiral theories, is fatal because the renormalizability of the theory is spoiled and unitarity is violated.[22] Since any realistic model must contain some chiral sector(s), there is no way of avoiding the presence of ASB. The only way out, then, is to make some ad hoc arrangements for canceling the anomalies.[23] This requirement also leads to severe restrictions on the choice of spacetime dimensions (10 or 26) and of symmetry groups ($SO(32)$ or $E8 \times E8$) in the context of superstring theories (cf. Green et al., 1987). While it is debatable whether such restrictions are a great success or a crushing defeat, there is no doubt that the investigation of ASB occupies a central place in the research on the foundations of QFT.

If the symmetries concerned are global, then the occurrence of ASB is harmless or even desirable, as in the case of global γ_5 invariance for explaining $\pi^0 \rightarrow \gamma\gamma$ decay (cf. Bell and Jackiw, 1969), or of scale invariance in QCD with massless quarks for obtaining massive hadrons as bound states. But the implications of the anomalous symmetry breakdown of scale invariance are extremely profound so that they demand separate discussions.

3.3. *Scale Invariance and Renormalization Group Approach*

The idea of scale dependence within the framework of QFT appeared earlier than the idea of scale invariance, and it can be traced to Dyson's work on the smoothed interaction representation (1951). In this representation, the low frequency part of the interaction can be treated separately from the high frequency part, which was thought to be ineffective, except in producing renormalization effects. To this end Dyson defined, adopting the guidelines of the adiabatic hypothesis, a smoothly varying charge of the electron and a smoothly varying interaction with the help of a smoothly varying parameter g. He then argued that when g is varied, some modification had to be made in the definition of the g-dependent interaction, in order to compensate for the effect caused by the change of the g-dependent charge. In line with this idea of Dyson, Landau and his collaborators developed a similar concept of smeared out interaction in a series of influential papers (Landau et al., 1954a, 1954b, 1954c, 1954d, 1956). In accordance with this concept, the magnitude of the interaction should be regarded not as a constant but as a function of the radius of interaction that must fall off rapidly when the momentum exceeds a critical value $P \sim 1/a$, where a is the range of the interaction. As a decreases, all the physical results tend to finite limits. Correspondingly, the electron's charge must be regarded as an as yet unknown function of the radius of interaction. With the help of this concept, Landau studied the short distance behavior of QED and obtained some significant results, which were referred to in the last section. Both Dyson and Landau had the idea that the parameter corresponding to the charge of the electron was scale-dependent. In addition, Dyson hinted, though only implicitly, that the physics of QED should be scale-independent; Landau, more explicitly, suggested that the interactions in QED might be asymptotically scale-invariant.

In later works, in particular those of Stueckelberg and Petermann (1953) and of Gell-Mann and Low (1954), Dyson's varying parameter g and Landau's range of interaction were further specified as the sliding renormalization scale, or subtraction point. In Gell-Mann and Low, the scale-dependent character of parameters and the connection between parameters at different renormalization scales were elaborated in terms of renormalization group transformations, and the scale-independent character of the physics was embodied in renormalization group equations. However, these elaborations were not appreciated until much later – the late 1960s and early 1970s – when a deeper understanding of the ideas of scale invariance and of renormalization group equations was gained, mainly through the researches of K. G. Wilson, the result of fruitful interactions between QFT and statistical physics.

The idea of the scale invariance of a theory is more complicated and very different from the idea of the independence of the physics on the renormalization scale as expressed by the renormalization group equations. The scale invariance of a theory refers to its invariance under the group of scale transformations. The latter are only defined for dynamical variables (the fields), but not for the dimensional parameters, such as masses, for otherwise a scale transformation would result in a different physical theory. While the physics should be independent of the choice of the renormalization scale, a theory may not be scale-invariant if there are any dimensional parameters.

In Gell-Mann and Low's treatment of the short distance behavior of QED, the theory is not scale-invariant when the electric charge is renormalized in terms of its value at very large distances. The scale invariance would be expected in this case because the electron mass can be neglected and there seems to be no other dimensional parameter appearing in the theory. The reason for the unexpected failure of scale invariance is due entirely to the necessity for charge renormalization: there is a singularity when the electron mass goes to zero. However, when the electric charge is renormalized at a relevant energy scale by introducing a sliding renormalization scale to suppress effectively irrelevant low energy degrees of freedom, there seems to occur an asymptotic scale invariance. This 'asymptotic scale invariance' is expressed by Gell-Mann and Low in terms of a scaling law for the effective charge and by the eigenvalue condition for the bare charge, that is, by the statement that there is a "fixed" value for the bare charge independent of the value of the measured charge.[24]

Although there was a suggestion by Johnson (1961) in the early 1960s that the Thirring model might be scale-invariant, the real advance in understanding the nature of scale invariance was made in the mid 1960s as a result of developments in statistical physics. Research in this area was also stimulated by the discovery of field-theoretical anomalies in the study of current algebra and in the short distance expansion of products of quantum field operators. Here we want to emphasize that the interaction between QFT and statistical mechanics, which can readily be discerned in the shaping of Wilson's ideas, played an important role in the development. Conceptually, the interaction is very interesting, but also quite complicated. In 1965, Widom (1965a, 1965b) proposed a scaling law for the equation of state near the critical point that generalized earlier results obtained by Essam and Fisher (1963) and Fisher (1964) concerning the relations among the critical exponents. Wilson was puzzled by Widom's work because it lacked a theoretical justification. Wilson was familiar with Gell-Mann and Low's work. Moreover, he had just found a natural basis for the renormalization group analysis, while working to develop a lattice field theory, by solving and eliminating one momentum scale for the problem (Wilson, 1965). At the time, Wilson realized that there should be applications of Gell-Mann and Low's idea to critical phenomena. One year later, Kadanoff (1966) derived Widom's scaling law using the idea – which essentially embodied the renormalization group transformation – that the critical point becomes a fixed point of the transformations on the scale-dependent parameters. Wilson quickly assimilated Kadanoff's idea and amalgamated it into his thinking about field theories and critical phenomena, exploiting the concept of broken scale invariance.[25]

Wilson had also done some seminal work in 1964 (unpublished) on operator product expansions (OPE), but had failed in the strong coupling domain. After thinking about the implications of the scaling theory of Widom and Kadanoff when applied to QFT, and after having investigated the consequences of Johnson's (1961) suggestion concerning the scale invariance of the Thirring model and that of Mack (1968) concerning the scale invariance of the strong interactions at short distances, Wilson reformulated his theory of OPE basing it on the new idea of scale invariance (1969). He found that QFT might be scale-invariant at short distances if the scale dimensions of the field operators, which are defined by the requirement that the canonical commutation relations are scale invariant, were treated as new degrees of freedom.[26] These scale dimensions can be changed by the interactions between the fields and can acquire anomalous values,[27] which mathematically correspond to the nontrivial exponents in critical phenomena.

The most important implications for the foundational transformations of QFT stemming from the dramatic advances in statistical physics can be summarized by two concepts Wilson stressed: (i) the statistical continuum limit of a local theory, and (ii) the fixed points of renormalization group transformations.

First, Wilson noticed that systems described by statistical physics and by QFT embodied various scales. If functions of a continuous variable, such as the electric field defined on spacetime, are themselves independent variables and assumed to form a continuum so that functional integrals and derivatives can be defined, then one can define a statistical continuum limit that is characterized by the absence of a characteristic scale. This means that fluctuations in all scales are coupled to each other and make equal contributions to a process. In QED calculations this typically leads to logarithmic divergences. Thus renormalization is necessary for the study of these systems. That this concept of statistical continuum limit occupies a central position in Wilson's thinking on QFT in general, and on renormalization in particular, is reflected in his claim that "the worst feature of the standard renormalization procedure is that it gives no insight into the physics of the statistical continuum limit" (1975, p. 775).

Second, the various parameters characterizing the physics at various renormalization scales reflect the scale dependence of the renormalization effects. These parameters are related to each other by the renormalization group transformations that are described by the renormalization group equations. In this sense, the renormalization group equations study the high energy behavior of QFT by following the variation of the effective parameters of the theory caused by the anomalous breakdown of scale invariance of the theory.

Since the late 1960s it has been recognized that the scale invariance of any quantum field theory is unavoidably broken anomalously because of the necessity of renormalization. This insight was foreshadowed by Adler (1969), Bell and Jackiw (1969), and others in 1969 in their studies of current algebra. They found that in perturbation theory the equal-time commutators of field operators were affected by renormalization. This discovery helped ascertain the existence of chiral anomalies and the anomalous dimension of quantum fields, and led to the idea of ASB.

A more convincing argument is based on the concept of dimensional transmutation. The scale invariance of a theory is equivalent to the conservation of the scale current in the theory. To define the scale current, a renormalization procedure is required, for, as a product of two operators at a same point, the scale current implicitly contains an ultraviolet singularity. However, even in a theory without any dimensional parameter, it is still necessary to introduce a dimensional parameter as a subtraction point when renormalizing, in order to avoid the infrared divergences and in order to define the coupling constant. And this breaks the scale invariance of the theory. The necessity to introduce a dimensional parameter was called "dimensional transmutation" by Coleman and Weinberg (1973). Precisely because of dimensional transmutation, the scale invariance in a renormalized theory is unavoidably broken anomalously, though the effects of this breakdown can be taken care of by the renormalization group equations.

In statistical physics, the renormalization group approach effects connections between physics at different scale levels. By scaling out the irrelevant short-range correlations and by locating stable infrared fixed points, it has made possible the conceptual unification of various descriptions – such as those of elementary excitations (quasi-particles) and collective ones (phonons, plasmons, spin-waves) – the explanation of the universality of various critical behavior, and the calculation of order parameters and critical components. In QFT, the same approach can be used to suppress the irrelevant low energy degrees of freedom, and to find a stable ultraviolet fixed point. In both cases, the essence of the approach, as Weinberg (1981) has indicated, is to concentrate on the relevant degrees of freedom for a particular problem,[28] and the goal is to find fixed point solutions of the renormalization group equations.[29]

According to Wilson, the fixed point in QFT is just a generalization of Gell-Mann and Low's eigenvalue condition for the bare charge in QED. At the fixed point a scaling law holds, either in Gell-Mann–Low–Wilson's sense or in Bjorken's, and the theory is asymptotically scale-invariant. The scale invariance is broken at nonfixed points, and the breakdown can be traced by the renormalization group equations. It is clear that if the renormalization group equations of a given field theory possess a stable ultraviolet fixed point solution, then in that field theory the high energy behavior causes no trouble, and it can thus be called, according to Weinberg (1978), an "asymptotically safe theory". An asymptotically safe theory may be a renormalizable theory if the fixed point it possesses is the Gaussian fixed point.[30] Weinberg, however, argued, and supported his position with a concrete example of a five-dimensional scalar theory, that the concept of 'asymptotic safety' is more general than the concept of renormalizability, and thus can explain and even replace it. There may in fact be cases in which theories are asymptotically safe but not renormalizable in the usual sense, if they are associated with a Wilson–Fisher fixed point.

The conceptual developments described in this section can be summarized as follows: in systems with many scales that are coupled to each other and without a characteristic scale, such as those described by QFT, the scale invariance is always anomalously broken due to the necessity of renormalization. This breakdown manifests itself in the anomalous scale dimensions of fields in the framework of OPE, or in the variation of parameters at different renormalization scales that is charted by the renormalization equations. If these equations have no fixed point solution, then they are not asymptotically scale-invariant and the theory is, rigorously speaking, nonrenormalizable;[31] if they possess a fixed point solution, then they are asymptotically scale-invariant and the theory is asymptotically safe. If the fixed point is Gaussian, then the theory is renormalizable. But there may be some asymptotically safe theories that are nonrenormalizable if the fixed point they possess is a Wilson–Fisher fixed point. With the occurrence of the more fundamental guiding principle of asymptotic safety, which is one of the consequences of the renormalization group approach, the fundamentality of the renormalizability principle began to be seriously challenged.

3.4. *Decoupling Theorem and Effective Field Theories*

According to the renormalization group approach, different renormalization prescriptions only lead to different parameterizations of a theory. An important application of this freedom in choosing a convenient renormalization prescription is embodied in the decoupling theorem, first formulated by Symanzik (1973), and then by Appelquist and Carrazzone (1975). The theorem is concerned with renormalizable theories in which some fields have masses much larger than the others and is based on power-counting arguments. The theorem states that in such theories a renormalization prescription can be found such that the heavy particles can be shown to decouple from the low energy physics, except for producing renormalization effects and corrections that are suppressed by a power of the experimental momentum divided by a heavy mass.

An important corollary to this theorem is that the low energy physics is describable by an effective field theory (EFT), which incorporates only those particles that are actually important at the energy being studied: there is no need to solve the complete theory describing all the light and the heavy particles (cf. Weinberg, 1980b). The EFT can be obtained by deleting all heavy fields from the complete renormalizable theory and suitably redefining the coupling constants, masses, and the scale of the Green's functions, using the renormalization group equations. Clearly, a description of the physics by an EFT is context-dependent. It is delimited by the experimental energy available, and thus able to keep close track of the experimental situation. This context dependence of an EFT is embodied in an effective cutoff that is represented by a heavy mass associated with SSB. Thus, with the decoupling theorem and the concept of EFT emerges a hierarchical picture of nature offered by QFT, one that explains why the description at any one level is so stable and is not disturbed by whatever happens at higher energies, and thus justifies the use of such descriptions.

There seems to be an apparent contradiction between the idea underlying the renormalization group approach and the idea underlying EFT. While the former is predicated on the absence of a characteristic scale in the system under consideration, the latter takes seriously the mass scale of the heavy particles, in which mass scale plays the role of a physical cutoff or a characteristic scale in the low energy physics that involves only the light particles. The contradiction disappears immediately, however, if we remember that the heavy particles still make contributions to the renormalization effects in EFT. Thus the mass scale of the heavy particles in an EFT actually plays only the role of a pseudocharacteristic scale, not a genuine one. The existence of such pseudocharacteristic scales reflects a hierarchical ordering of couplings at different energy scales, but it does not change the essential feature of systems described by QFT, namely, the absence of a characteristic scale and the coupling of fluctuations at various energy scales. While some couplings between fluctuations at high and low energy scales exist universally and manifest themselves in the renormalization effects in low energy physics, others are suppressed and reveal no observable clues in low energy physics.

The above assertion that the decoupling is not absolute is reinforced by the important observation that the influence of the heavy particles on the low energy physics *is* directly detectable in some circumstances:

if there are processes (e.g., those of weak interactions) that are exactly forbidden by symmetries (e.g., parity, strangeness conservation, etc.) in the absence of the heavy particles that are involved in symmetry-breaking interactions that lead to these processes (e.g., *W*- and *Z*-bosons in the weak interactions), then the influence of the heavy particles on the low energy phenomena are observable, although, due to the decoupling theorem, suppressed by a power of energy divided by the heavy mass. Typically, these effects can be described by an effective nonrenormalizable theory (e.g., Fermi's theory of weak interactions), which, as a low energy approximation to a renormalizable theory (e.g., the electroweak theory), possesses a physical cutoff or characteristic energy scale set by the heavy particles (e.g., 300 Gev for the Fermi theory). When the experimental energy approaches the cutoff energy, the nonrenormalizable theory becomes inapplicable, and new physics appears that requires for its description either (i) a renormalizable theory or (ii) a new, effective theory with a higher cutoff energy. The first choice represents the orthodoxy. The second choice presents a serious challenge to the fundamentality of the renormalizability principle, and is presently gaining momentum and popularity.

3.5. *A Challenge to Renormalizability*

The concept of an EFT clarifies how QFT at different scales takes different forms and allows two different ways to look at the situation:

(i) If a renormalizable theory at high energy is available, then the effective theory at any lower energy can be obtained in a totally systematic way by integrating out the heavy fields of the theory. In this way, the renormalizable electroweak theory and QCD can be understood as effective theories at low energy of some grand unified theory. They thus lose their presumed status of being fundamental theories. Another possibility, also compatible with this way of looking at the situation, is to assume that there exists a tower of effective theories that contains nonrenormalizable interactions, each with fewer numbers of particles and with more small nonrenormalizable interaction terms than the previous. When the physical cutoff (heavy particle mass *M*) is much larger than the experimental energy *E*, the effective theory is approximately renormalizable, the nonrenormalizable terms being suppressed by a power of E/M.

(ii) The second way corresponds more closely to what high energy theorists actually do when studying physics. Since nobody knows what the renormalizable theory at the unattainable higher energies is, or even whether it exists at all, we have to probe the accessible low energy first and design representations that fit this energy range. We then extend our theory to higher energies only when it becomes relevant to our understanding of physics. We thus obtain an endless tower of theories,[32] in which each theory is a particular response to a particular experimental situation and none can ultimately be regarded as the fundamental theory. In this approach, the requirement of renormalizability can be replaced by a condition on the nonrenormalizable interactions in the effective theories: all the nonrenormalizable interactions in an effective theory describing physics at a scale m must be produced by heavy particles with a mass scale M ($\gg m$), and are thus suppressed by powers of m/M. Furthermore, in the renormalizable effective theory including the heavy particles with mass M, these nonrenormalizable interactions must disappear.

These clarifications, together with the renormalization group equations, have helped physicists to come to a new understanding of renormalization. As David Gross put it, renormalization "is an expression of the variation of the structure of physical interactions with changes in the scale of the phenomena being probed" (1985, p. 153). Notice that this new understanding is very different from the old one, which focused exclusively on the high energy behavior and on ways of circumventing the divergences. It shows a more general concern with the finite variations of the various physical interactions with finite changes of energy scales, and thus provides enough leeway for considering nonrenormalizable interactions.

A significant change in the attitude of physicists with respect to what should be taken as guiding principles in theory construction has taken place in recent years in the context of the development of EFT. On the heels of Dyson's classic work, renormalizability was taken for many years as a necessary requirement for acceptable quantum field theories.

Now, reflecting the fact that experiments can probe only a limited range of energies, it seems natural to take EFT as a general framework for analyzing experimental results. Since nonrenormalizable interactions occur quite naturally within this framework, there is no a priori reason to exclude them when constructing theoretical models to describe currently accessible physics.

In addition to being compatible and congenial with the new understanding of renormalization, taking nonrenormalizable interactions seriously is also supported by some other arguments. First, nonrenormalizable theories are malleable enough to accommodate experiments and observations, especially in the area of gravity. Second, they possess predictive power and are able to improve this power by taking higher and higher cutoffs. Third, because of their phenomenological nature, they are conceptually simpler than the renormalizable theories, which, as stressed by Schwinger, involve physically extraneous speculations about the dynamic structure of the physical particles. The traditional argument against nonrenormalizable theories was that they are unstable (undefinable at energies higher than their physical cutoff). A more extended discussion of the question of the stability of such theories will be given in the next section. Here we focus on the high energy behavior of nonrenormalizable interactions, which many physicists have regarded as one of the most fundamental questions in QFT.

As noted above, within the original framework of EFT which takes renormalizability as its conceptual basis, nonrenormalizable theories as low energy approximations to renormalizable ones secure standings only as auxiliary devices. When the experimentally available energy approaches their cutoffs and new physics begins to appear, they become incorrect and have to be replaced by renormalizable theories. Within the framework of Weinberg's asymptotically safe theories, nonrenormalizable theories have acquired a more fundamental status. Nevertheless, they still share a common feature with EFT, namely, all the discussion of them is based on taking the cutoff to infinity, and thus falls into the category of formalistic interpretations of the cutoff.

However, if we take the idea underlying EFT to its logical conclusion, then a radical change in outlook takes place and a new perspective appears; a new interpretation of QFT can be developed and a new theoretical structure of QFT waits to be explored. Thoroughgoing advocates of EFT, such as Georgi (1989b) and Lepage (1989), would argue that when the experimentally available energy approaches the cutoff of alizable effective theory, we can replace it with another nonrenormalizable effective theory of much higher cutoff. In this way, the high energy behavior of the nonrenormalizable interaction above the cutoff will be properly taken care of by:

(i) the variation of the renormalization effects, caused by the change of the cutoff and calculable by the renormalization group equations; and
(ii) additional nonrenormalizable counter terms.[33]

Thus, at any stage of development, the cutoff is always finite and can be given a realist interpretation, as was argued in Section 3.1. In addition to the finite cutoff, we also find two new ingredients that are absent or forbidden in the traditional structure of QFT but are legitimate and indispensable in the theoretical structure of the new formulation of QFT. These are the variations of renormalization effects with specific changes of cutoff and the nonrenormalizable counterterms that are legitimatized by the introduction of finite cutoff.

These foundational transformations in the conceptualization of renormalization, which stem partly from an internal conceptual evolution and were inspired to a large extent by the significant progress in statistical physics, have provided fertile soil for the acceptance and the further development of Schwinger's insightful ideas. These had been advanced in his criticism of renormalization theory and of the operator formulation of QFT, and were detailed in the presentation of his source theory. Schwinger's views strongly influenced Weinberg in his work on the phenomenological Lagrangian approach to chiral dynamics and on EFT. We can easily find three features shared by Schwinger's source theory and the new formulation of QFT:

(i) the denial that they are fundamental theories;
(ii) their flexibility in being able to incorporate new particles and new interactions into the existing schemes; and
(iii) the possibility of each of them to consider nonrenormalizable interactions.

However, a fundamental difference exists between the two schemes. The new formulation of QFT is still a local operator field theory and contains no characteristic scale, and thus has to deal with the contributions from fluctuations at arbitrarily high energy. On the other hand, Schwinger's theory is a thoroughly phenomenological one, in which the numerical field, which is different from the operator field, is only responsible for the one-particle excitation at low energy. There is thus no question of renormalization in Schwinger's theory. On the other hand, in the new formulation of QFT, renormalization has taken a more and more sophisticated form, has gained in predictive power, and has become an evermore powerful calculational tool.

4. PHILOSOPHICAL RAMIFICATIONS

Having examined both the cognitive content of renormalization theory and the transformations of its structure over the last four decades, we are in a better position to clarify the implications and ramifications of the theory from a philosophical perspective. Most notably, we found that the recent developments support a pluralism in theoretical ontology, an antifoundationalism in epistemology and an antireductionism in methodology. These implications are in sharp contrast with the neo-Platonism implicit in the traditional pursuit of quantum field theorists, which took mathematical entities as the ontological foundation of physical theories and which assumed that, through rational (mainly mathematical) human activities, one could arrive at an ultimate stable theory of everything. Also, contrary to the previous image of scientific theories that was implicit in the mathematical structure of QFT, the new image fostered by the EFT approach is that scientific theories are not to be conceived as necessary products of scientific rationality, but rather should be seen as contingent descriptions of nature, revisable in the course of changing circumstances. These implications have important bearing on the debate about realism versus instrumentalism.

4.1. *Atomism and Pluralism in Theoretical Ontology*

The various models developed within the framework of QFT for describing the subatomic world are atomistic in nature: the particles described by the fields that appear in the Lagrangians are to be regarded as elementary constituents of the world. However, the atomism, or the notion of atomicity, adopted by unrenormalized theories have a halfway character. As clearly pointed out by Schwinger, the reason for this is that unrenormalized operator field theories contain an implicit assumption about the inner structure of physical particles that is sensitive to the details of dynamic processes at high energy. Mathematically, this assumption has manifested itself in the divergent integrals.[34] Metaphysically, the assumption implies that there exist more elementary constituents than the physical particles described by the fields appearing in a Lagrangian, and thus contradicts the status of the particles, or that of the fields, as basic building blocks of the world.

The fundamental significance of the renormalization procedure in this regard can be described in the following way: by removing any reference to inaccessible, very high energy domains and the related structural assumption, the renormalization procedure reinforces the atomistic commitment of QFT. This is because the particles or the fields appearing in the renormalized theories do act as the basic building blocks of the world. But note that atomicity here no longer refers to the exact point model. To the extent that one removes the reference to the inaccessible very high energy domain – which arises conceptually in the exact point model by virtue of the uncertainty principle – renormalization blurs any point-like character. This spatially extended yet (seemingly) structureless quasi-point model, adopted or produced by renormalized QFT, is justified in two ways. On the one hand, it is supported by its empirical success. It is also justified philosophically by arguing that, as long as the experimental energy is not high enough to detect the inner structure of the particles, so that all the statements about their (small distance) structure are essentially conjectures, the quasi-point model is not only an effective approximation for experimental purposes, but it also reflects a necessary stage of cognition that we must go through.

Compared to the 'true' theory that Dirac (1983) and Tomonaga (1965) once yearned for, and compared to the 'theory of everything' that superstring theorists are still searching for,[35] the quasi-point model seems to be merely a mathematical device needed and useful in a transition period. A disciple of Dirac would argue that the quasi-point model should be discarded once the structure of the elementary particles is known. This is true. But those committed to atomism would argue that in physics (as presently formulated and practiced) the structural analysis of objects at any level is always based on (seemingly) structureless objects – genuine or quasi-point-like – at the next level. For example, the analysis of deep inelastic lepton-hadron scattering would be impossible if there were no analysis of elastic scattering of the partons as a basis. Thus the adoption of the quasi-point model seems to be unavoidable in QFT. This is probably what Feynman meant when he stated that "[w]e will start by supposing that there are (quasi-point particles) because otherwise we would have no field theory at all" (1973, p. 775).

The dialectics of atomism in this context can be summarized as follows: on the one hand, the structureless character of particles as we know them at any level is not absolute but contingent and context-dependent, justified only by relatively low energy experimental probing. When the energy available in experiments becomes high enough, some of the inner structure of the particles sooner or later is revealed, and the notion of the absolute indivisibility turns out to be an illusion. On the other hand, with the unraveling of the structure of particles at one level, there emerge at the same time – as a precondition for the unraveling – (seemingly) structureless objects at the next level. And thus the original pattern of 'structured objects being expressed in terms of (seemingly) structureless objects' remains, and will remain as long as QFT remains the mode of representation. Thus the idea expressed by the bootstrap hypothesis in *S*-matrix theory that everything is divisible is incompatible with the atomistic paradigm adopted by QFT.

The EFT approach extends the atomistic paradigm further, and within that framework the domain under investigation is given a more discernible and a more sharply defined hierarchical structure. The hierarchy is delimited by mass scales associated with a chain of spontaneously broken symmetries and is justified by the decoupling theorem. The examination of the hierarchical structure from a metaphysical perspective yields two seemingly contradictory implications that are worth noting. On the one hand, the hierarchical structure seems to lend support to the possibility of interpreting physical phenomena in a reductionist or even reconstructionist fashion, at least to the extent that SSB works. Most of the efforts expended during the past two decades by the mainstream of the high energy physics community, from the standard model to superstring theories, can be regarded as exploring this possibility. On the other hand, taking the decoupling theorem and EFT seriously would entail considering the reductionist (and a fortiori the constructivist) program an illusion, and would lead to its rejection and to a point of view that accepts emergence, hence to a pluralist view of possible theoretical ontologies.[36]

The decoupling theorem does not reject the general idea of causal connections between different hierarchical levels. In fact, such connections are assumed to exist and to be describable by the renormalization group equations: they are thus built into the very conceptual basis of the theorem. It is the attempt to give the connections universal significance and the stipulation of their direct relevance to scientific inquiry that is rejected. More precisely, what is to be rejected is the suggestion that it is possible simply by means of these kinds of 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.* The necessity, as required by the decoupling theorem and EFT, of an empirical input into the theoretical ontologies applicable at the lower energy scales – scales to which the ontologies at the higher energy scales have no direct relevance in scientific investigations – is fostering a particular representation of the physical world. In this picture the latter can be considered as layered into quasi-autonomous domains, each layer having its own ontology and associated 'fundamental' laws. This hierarchical pluralism in theoretical ontology does not mean to reject the causal connections between ontologies at the different levels. So in a weak sense it still falls into the category of atomism: it merely rejects the possibility of deducing various entities from some basic ontology.

It is precisely the strong antireductionist commitment concerning the relationship between different hierarchical levels that differentiates the pluralistic version of atomism that is nurtured by EFT from the cruder version of atomism that is adopted by conventional QFT, the constitutive components of which are reductionism and constructionism. In addition, the emphasis on an empirical input that is historically contingent also sharply contrasts the hierarchically pluralistic version of atomism with the neo-Platonic mathematical atomism implicit in the traditional pursuit of quantum field theorists. The latter takes ahistorical mathematical entities for its ontological foundation and assumes that all empirical phenomena can be deduced from it. These distinctions in ontological commitment have direct relevance to the epistemological and methodological components of renormalization theory. We turn to a discussion of these matters in the next few sections.

4.2. *Antifoundationalism in Epistemology*

In our discussion of the foundational transformations of renormalization theory, we pointed out that the interactions between QFT and statistical physics played a crucial role. For example, the concept of SSB – which is indispensable both in the proof of the renormalizability of massive non-Abelian gauge theories and for delimiting the hierarchical structure of effective field theories – was first developed in the study of superconductivity. Similarly, the concept of broken scale invariance and the related renormalization group approach, which led to the EFT approach and to a new understanding of renormalization, were the result of the amalgamation of ideas and techniques developed in investigations of the short distance behavior of QED, with some arising in the study of critical phenomena. However, a question of fundamental significance was left unanswered: What is it that makes the exchange possible and so extremely fruitful?

At first sight the exchanges seem to be quite mysterious. Nor is the puzzle resolved by a historical analysis of the developments. Why are the physical insights obtained from one phenomenological domain (e.g., spins in crystal lattices) relevant, translatable, and applicable to another entirely different domain (e.g., continuous fields)? More specifically, even after rewriting the lattice formalism in a continuum language, which is mathematically possible, it is still very difficult to understand why the formalism of critical phenomena is applicable to QFT and vice versa. In particular, the effective Hamiltonians of statistical physics involve terms of arbitrary complexity and are thus quite different from the formalism of renormalizable field theories that involves only a finite number of interaction terms. Similarly, the concepts employed in the theory of critical phenomena (QFT), such as lattice spacings and block spins (coupling-constant and field renormalization) are designed to deal with infrared (ultraviolet) divergences, a problem that has little to do with QFT (critical phenomena).

Assuredly, the puzzle can be solved in a realist-essentialist way. The interactions between different physical theories can be explained by, or taken to argue for, the transcendental unity of physical phenomena on the ontological plane, and/or the universality of physical truths on the epistemological plane. One prominent example of a commitment to realism-essentialism is neo-Platonism, which frequently finds many adherents among mathematical physicists. In the last three decades, neo-Platonists working on fundamental physics, within the context of QFT, have taken ahistorical mathematical entities and relations, particularly gauge symmetries and supersymmetries, and their representations, as expressing true reality and/or manifesting the hidden essence existing beneath overt phenomena, both in terms of entities and their structural patterns. By revealing themselves in the real nature of various phenomena, they are appropriated to constitute the universal foundation of physical theories. According to this view, all physical (i.e., non-pure-mathematical) insights lose their fundamental importance and survive only as suggestive heuristics, and all theoretical activities in physics can be reduced to the search, by means of mathematical reasoning, for the one key factor (mathematical group structure) that would explain in a consistent fashion the complexity of the universe (cf. Radicati, 1984).

The latest example of such an overly grandiose and totalizing conception of physical theory is the search for the theory of everything by superstring theorists.[37] Although the pursuit has not been successful, it has deep roots that may be traced back to the ideas of Felix Klein (1872, 1918), Sophus Lie (Lie and Engel, 1893), Minkowski (1908, 1909), Weyl (1918a, 1918b, 1929), Cartan (1922), Einstein, and Heisenberg (1960, 1966). It is nurtured by "the unreasonable effectiveness of mathematics in the natural sciences", as Wigner once put it (1960). Among the spectacular successes of this approach we find Dirac's prediction of the positron (1930), Wigner's descriptions of atomic and nuclear spectroscopy (1931, 1937), Gell-Mann and Neéman's classifications of hadrons (1964), and Weinberg's and Salam's predictions of intermediate bosons and neutral currents. Thus it should not come as a surprise that at present a great many of the theorists in fundamental physics, directly under the sway of the successes of the Weinberg–Salam model, are preoccupied by this mathematical pursuit.[38]

As prevalent as this mathematical foundationalist tendency is, the recent developments in the theory of renormalization, from the renormalization group method to the EFT approach, have pointed in the opposite direction: the empirical input of theoretical ontologies in different domains of investigation have been emphasized and have provided a strong argument for the fundamental importance of phenomenological approaches. In contrast to a totalizing conception of physical theory, a phenomenological approach supports a localist view that characterizes physical (or more generally, scientific) theories as historically situated and context-dependent.[39] This view applies not only to the phenomenological laws describing natural phenomena, but also to the fundamental laws that explain phenomena and give them meaning.[40] By accepting the existence of fundamental laws, the localist view distances itself from another foundationalist tendency that takes sensory experience as the foundation of human knowledge.

The application of a localist view to fundamental laws needs to be explicated and a philosophical justification of the application has to be provided. As we have seen, the localist view of theory stems from the difficulties encountered in QFT with infinities, and finds its justification in an empiricist position regarding knowledge of the natural world, according to which knowledge is constitutive rather than given. Localists believe that our knowledge of the natural world is constructed through a complex feedback process in which our sensory input and our theoretical models are assimilated and accommodated in a self-modifying sequence of learning, hypothesizing, deducing, predicting, testing, and correcting the entire process being constrained by spatio-temporal reality. The limited nature of our experience in producing knowledge of the world undermines the universal claim of physical laws: it only allows ascertaining family resemblance (regularities) in local region of space and time. From local regularities we cannot construct physical theories that are unique and necessary. On the contrary, all theories are context-dependent, culturally relative, and historically changeable.

Furthermore, the concept of fundamental laws which explain phenomena and give them meaning in terms of abstract concepts raises three questions that need to be answered:

(i) What makes abstract concepts possible within the localist framework?
(ii) What constitutes giving meaning to phenomena?
(iii) What constitutes scientific explanation?

The answer to (i) is shifting metaphors. Metaphors connect the most abstract of concepts with the reality of everyday life via a kind of similarity in structures, thus ascribing meaning to abstract concepts. According to this view, the meaning of a concept is metaphorical in the typical instance, and literal (truth-functional) only in the limiting case. Thus, all theoretical concepts are dynamic in the sense that they are subject to metaphorical expansion and transformation. Concerning (ii), localists believe that due to the feedback nature of the learning process, meaning cannot be constituted by any atomistic fact or reference, as claimed by Russell (1914) and the early Wittgenstein (1922), but rather is conferred by holistic networks, and their dispositions and transformations. And since the distinct metaphorical contexts that define the intersections and interactions of different parts of the network do so differently, no phenomena can have fixed meaning across different metaphorical contexts. This means that all understanding of phenomena requires the assimilation of a dynamic network of meanings that is controlled and operated by metaphorical associations and metaphorical shifts. Thus, a provisional answer to (iii) is that a scientific explanation can be understood as a metaphorical redescription of the domain of phenomena.[41]

The metaphoric view of scientific explanation has created a new potential for understanding the interactions between different physical theories. Roughly speaking, the exchanges manifest themselves either in mathematical analogies or in physical analogies. A mathematical analogy, such as renormalization group equations and fixed points in statistical physics and in QFT, works principally because different physical interpretations of the mathematical formalisms in different domains of phenomena are connected by metaphorical transformations of concepts involved in the formalisms. A physical analogy, such as the suppression of irrelevant degrees of freedom by using renormalization group equations both in QFT and in statistical physics, is itself a metaphorical expansion of the physical insights obtained in one domain of phenomena to apply to another.

The epistemological position supported by the recent developments in renormalization theory, especially by the EFT approach, is empiricist in nature. Physical theories are justified by empirical data from which the theories are abstracted. They are effective instruments for organizing the data by imposing local order and coherence, and they conceive and express local causal regularities. Physical theories can be seminal and suggestive in a domain beyond the boundary delimited by the data because the local regularities are expressed in an apparently universal mathematical formalism, and are thus able to be metaphorically expanded in a coherent way. But the extrapolation cannot be fully justified by the mathematical formalism itself and awaits an empirical justification. It seems to us that what is most significant in these recent developments is the rejection by the EFT approach of the claim of universality for the mathematical formalism in QFT, and its emphasis on the local nature of theoretical structures and on their variation with changes in the scale of the phenomena being probed.

This position rejects uncompromisingly the idea successively advanced during the last fifteen years by grand unified theorists, supergravity theorists, and superstring theorists that the development of fundamental physics will end with the discovery of an ultimate, definitive, and conclusive mathematical formalism. Rather, the development is taken as a process of successive extrapolations that is assumed not to have an end, with every step of the extrapolation being justified by a collective reinterpretation of theory and observation before and after the extrapolation, and through debates within the physics community about the meaning of theoretical concepts. It is also taken to be a process of social construction in which the ahistorical quest for logical coherence and stability is but one component, and even this component has to be defined historically because the understanding of coherence and stability differs from time to time.[42] The social character of the process is highlighted by the fact that it also incorporates an ideological component expressed by cultural preference[43] and aesthetic considerations, examples of which are the appeal to unification and the criteria of simplicity. Most important, the process is characterized by a socially defined instrumentalist progress: physics becomes more powerful and is judged to be so by the pragmatic criterion of empirical adequacy in an ever expanding local domain (success in explanation, prediction, and control). However, this instrumentalist progress neither presupposes nor implies any universal or permanent truth for theories that extrapolate far from the data.

In summary, the empiricist position in epistemology that is supported by the recent developments in renormalization theory is characterized by its antiessentialism and its antifoundationalism, its rejection of a fixed underlying natural ontology expressed by mathematical entities, and its denial of universal, purely mathematical truths in the physical world. This position is also characterized by its constructivism, emphasizing the socially constructive nature of physical theories, their local character, and their dynamics realized through a social process of metaphorical expansion and transformations.

4.3. *A Reappraisal of Criteria for Theory Acceptance*

QFT, as other branches of mathematical physics, aims to find laws of nature that are fundamental. In contradistinction to phenomenological laws that can be induced from a limited number of observations or experiments, fundamental laws can (usually) only be formulated by working within a mathematical framework and are frequently suggested by thought processes in which unobservables are introduced. The particular aim of mathematical physics has determined its specific procedures to be hypothetico-deductive. In this approach, hypotheses, including idealized models, do not (necessarily) originate in experiments. Most often, they are directly inspired by mathematical symbolism and are suggested by symbolic reasoning, as in the case of Dirac and Gell-Mann,[44] or simply come from the free creation of the mind, as in the case of Einstein (cf. 1936). They are able to generate new ideas that encompass notions far beyond deductions from experiment. This fact, however, does not mean that hypotheses in mathematical physics are autonomous and are not constrained by the external world. On the contrary, they have to undergo an external test in order to establish their validity. Since direct tests for highly abstract hypotheses are impossible, mathematical physicists have to deduce certain factual consequences of their hypotheses for their confirmation. But no particular hypothesis can be thought of as being unambiguously refuted or confirmed – as Duhem (1954), Milne (1929), and Hesse (1974) have forcefully argued – by virtue of the network nature of physical theories in which a hypothesis is only one thread. For this reason, some refinements of the hypothetico-deductive method are required: criteria have to be formulated so that theories constructed with this method can be judged acceptable or unacceptable,[45] and a rational choice can be made should there be several acceptable theories.

Paramount among the criteria for theory acceptance are the empirical adequacy and the stability of the theory. Dyson's renormalizability requirement was soon elevated to be a regulative principle because the renormalization procedure was thought by some physicists to have made QFT (i) calculable, predictive, and empirically adequate, and (ii) theoretically stable. However, as we have indicated in Section 2, later inquiries showed that renormalizable theories, although they did enjoy great success in their empirical adequacy, were far from being stable. It is clear from this example that three questions can readily be asked:

(i) What is the precise meaning of stability?
(ii) What is the relationship between adequacy and stability?
(iii) What is the primary determinant of theory appraisal? Is it adequacy, stability, or some other criterion?

Let us take the third question first.[46] Among the criteria for theory appraisal, those that are based on ideology, such as cultural preference,[47] aesthetic considerations (e.g., unification, simplicity), and plausibility (e.g., non-ad-hoc-ness and naturalness), possess only heuristic value without logical compulsion.[48] Those grounded in (nonempirical) metaphysics, such as constitutive metaphysics (i.e., unexplained explainers, their qualities and modes of interactions), and that prefer explanation types consilient with accepted ideas, assign the metaphysical intelligibility and explanatory value to the theories under consideration. Those from the constitutive level (with respect to theory, not to the object of knowledge), such as empirical adequacy and stability in theorizing, are thought to be necessary at least for the final acceptance of a theory. The actual appraisal of a scientific theory is a complicated and intricate business that takes place on several levels simultaneously. Weakness on one level may be compensated for or neutralized by strengths at another level, and no single criterion can be taken as the universally valid determinant in the appraisal of scientific theories. In the appraisal of a particular theory by a scientific community, the determination of which criterion should act as the determinant and which criteria should recede into the background is more than a matter of logic and depends on the concrete situation, the historical background, and the cultural context.[49] Leibniz's and Berkeley's critiques on the metaphysical level of Newton's concept of space without substance[50] ceased to bother natural philosophers in the eighteenth century because of the empirical success that Newtonian mechanics enjoyed.[51] The same critiques, however, became a crucial consideration in Mach's and Einstein's assessments of Newtonian mechanics despite its empirical successes. One of the factors that contributed to the new weight of metaphysical consideration was the general crisis faced by the mechanical world view toward the end of the nineteenth century. Another illuminating example is the following: grand unified theories (GUTs) have been retained by the high energy physics community on rational grounds in the face of contrary empirical evidence (the seeming failure to find any event of proton decay despite intensive search) due to the prevailing faith in renormalizable theories.[52]

It is widely held that the renormalization procedure to remove infinities has established the stability of QFT. Another widely accepted idea is that nonrenormalizable theories are unacceptable because of their instability, even though they are empirically adequate in the energy range accessible to us. Both ideas presume a global and universal conception of stability that refers to all energy ranges and to all possible kinds of interactions, respectively. But it would be too narrow-minded to exclude nonrenormalizable theories from consideration, while important interactions such as gravity steadfastly resist description by renormalizable theories. Also, it would be expecting too much to assume renormalizable theories to be stable in this global and universal sense, while ignoring the inconsistency of these theories (as pointed out in Section 2) and the incomprehensibility of some features of these theories in terms of accepted ideas. We have in mind such things as the decoupled ghost states and the cancelation of infinities, which presume the physical reality of ghost states and infinite totalities. A proper assessment of renormalizable theories and nonrenormalizable theories requires a clarification of the concept of stability, and this leads us to the first question.

In mathematical physics, the concept of stability comprises three related but distinguishable components, namely, logical consistency, mathematical definability and decidability, and conceptual stability. Defined as a requirement of noncontradiction, logical consistency is certainly necessary in formal thinking. Yet mathematical physics is more than a formal system. Its understanding requires a semantic and a pragmatic analysis of the concepts employed in the system, and this complicates the situation. For the sake of distinction, we shall refer to logical consistency with all the semantic and pragmatic complications as *conceptual stability*. One complication that is introduced by the requirement of conceptual stability is the following: conceptual stability always involves the requirement that a new concept and its implications be consilient with the network of accepted ideas in the system. In this sense, the conceptual stability is not a necessary criterion for theory acceptance. For example, Newton's concept of action-at-a-distance is not consilient with the idea of contact interactions that people had acquired from everyday experience, and the concept of quantum jumps is not consilient with the spatiotemporal picture of physical events that had underlain the classical conception of nature, although in both cases mathematical stability, which lies between the logical and the conceptual components of the stability, is not disturbed. That conceptual stability is not at the heart of mathematical physics can also be justified from an epistemological point of view: the metaphoric nature of meaning and scientific explanation (see Section 3.2) entails that conceptual systems are dynamic in nature, subject to metaphoric shifts, associations, and transformations, and obey rules different from those of formal logic. The importance of conceptual consilience recedes while the metaphoric network tries to assimilate apparent contradictions. This explains why at present few people pay serious attention to the once paradoxical situation of wave-particle duality in quantum physics.

The mathematical component of the concept of stability needs more explanation. The reason why this component is at the core of the concept in mathematical physics is not only that the theoretical concepts and hypotheses are expressed in terms of a mathematical formalism but also that the logical validity of the conceptual system is realized through a stable formalism. Additionally, and of greater importance to our concern here, is the fact that the universality of theoretical claims and the global character of conceptual stability are largely derived from and justified by the formal character of the symbolic system of mathematics that does not refer to any concrete meaning. For example, the most striking and persistent symptom of the conceptual instability of QFT is the occurrence of infinite integrals, indicating the lack of definability of QFT at high energy and thus challenging its global character. Traditionally, a limitation to the global validity of a mathematical formulation is taken as a decisive criterion for its rejection. Illustrative examples include Fermi's theory of weak interactions and models that contain chiral and/or gravitational anomalies. A serious challenge to this negative attitude toward the limits of the globalism and universality of mathematical formalism comes from the EFT approach. It takes the limit positively as indicating the hierarchical structure of the phenomena under investigation, and as delineating the valid domain of an effective Lagrangian. This challenge to the traditional understanding of the concept of mathematical stability requires justification and invites a new understanding of the concept. However, before presenting an exposition of the new understanding, some preliminary comments are in order.

The connection between the problem of mathematical stability and the existence of infinite quantities in the context of QFT is not peculiar to QFT. Rather, it is universal and the whole problem of mathematical stability actually originates from attempts to understand properly and to treat the infinite totalities that occur in mathematics. If mathematics is to be restricted to the description of observable objects, which are always finite in number and size, then no antinomies generated from infinite totalities would arise. However, unavoidably adjoined to observable objects are theoretical objects (various infinite totalities), to which there correspond in mathematics ideal concepts, statements, and inferences involving infinite totalities. The introduction of infinite totalities is inevitable even in classical mathematics. For example, the central notion of classical analysis, that of a real number, is defined in terms of actual infinite totalities. The situation became much more complicated when transfinite mathematics was first introduced by Cantor, and then incorporated by Whitehead and Russell into their *Principia Mathematica* (1910–13). It was quickly discovered that transfinite mathematics led to contradictions – such as "the class of all cardinal numbers" and "the class of all those classes which do not contain themselves as member", etc. Three responses were generated:

(i) The logicist response by Russell and others was to devise ad hoc remedies for avoiding antinomies while accepting the concept of actual infinite totalities.
(ii) The intuitionist response by Brouwer (1913) and others was to eliminate the notion of actual infinite totalities from mathematics.
(iii) The formalist response by Hilbert (1918, 1930) and others was to try to accommodate transfinite mathematics within a mathematics that is conceived as concerned with observable objects.

Hilbert's methodological transfinite response is the most relevant one to our concern with the stability problem. According to Hilbert's formalism, transfinite concepts are admitted into mathematical theories only because of their usefulness for such purposes as simplification and unification. No full ontological status is going to be accorded to them. The amplification of the mathematical system can only be justified by proving its stability.[53] Thus the concept of stability became the cornerstone of Hilbert's formalist program of reconstructing the whole edifice of mathematics. In particular, this concept is crucial in order to legitimate the use of ideal concepts involving infinities. Any unstable system is regarded by the formalist-finitist as meaningless and has to be rejected. This attitude, which originated in the formalist-finitist philosophy of mathematics, later became very popular and continues to haunt mathematical physicists, especially axiomatic-constructive field theorists.

However, the collapse of the formalist-finitist program in mathematics, in the wake of Gödel's work, has suggested an alternative attitude toward mathematical stability. The new understanding takes ordinary mathematical experience seriously and acknowledges that in formal systems there are parts of the mathematical practice in which we encounter statements decidable by a finitist method within the system, but that there are also statements whose truth value can be convincingly and consistently decided, but usually only by semantic analysis, or by other external, informal, and empirical considerations. A supporting example is the suggestion by Gödel (1938, 1939, 1940, 1947), which was later proven by P. J. Cohen (1963), that the continuum hypothesis is formally undecidable within axiomatic set theory. Thus the meaning of the hypothesis is independent of formal axioms, and the stability of axiomatic set theory loses its formal character. The stability is established by semantic analysis, which is certainly not as pure as the formal approach, and depends on context and interpretation (cf. Kreisel, 1976, 1980).

Thus, according to this new understanding, even in pure mathematics there is no sharp boundary line separating mathematical stability, which is pure, formal, and universal, and conceptual stability, which is impure, context-dependent, local, and subject to revisions with changes in the context. Notice that we are not suggesting a thoroughgoing intuitionist concept of constructibility in mathematical practice that rejects the significance of stability in nonconstructive mathematical practice. Rather, we want to emphazise that both mathematical and conceptual stabilities are committed to certain local characteristics.

We are now in a position to answer the second question concerning the relationship between stability in theorizing and empirical adequacy. Implicit in the question as to why a globally inconsistent theory, such as QFT, can be empirically adequate is a global view of stability that originated in the failed formalist-finitist program. However, our legitimation of a nonglobal view has provided room for a globally unstable theory to be empirically adequate, and only required that local stability in theorizing is necessary – though not sufficient – for empirical adequacy, the latter always being local in character. We have thus answered the question.

The priority and dominance of local adequacy over global stability is in general agreement with the methodological implications of the recent developments in renormalization theory. It implies a rejection of any attempt to mathematize physics totally – a tendency that originated with Descartes, Leibniz, and Eddington, and which has led to striving for global stability, universality, and necessity, to trying to derive all the observational facts of nature from hypothetic overarching principles, and, more concretely, to the formulation of the theories of supergravity and superstrings. While committed to the general framework of the hypothetico-deductive method, local adequacy also implies stressing the indispensability of phenomenological approaches and emphasizing their increasing importance. Phenomenological approaches are inductive in nature. They pay more attention to the empirical input from observations and experiments. They extrapolate theories from a local domain to a larger but still local domain, and verify that every step of the extrapolation is empirically confirmed.[54]

4.4. *Realism versus Instrumentalism*

Steven Weinberg in his Nobel lecture stated that the constraint of renormalizability "may point the way to the one true theory" and that "renormalizability might be the key criterion which would help us to pick out the one true physical theory of the infinite variety of conceivable quantum field theories" (1980a, p. 516). His strong belief in the existence of a "one true theory" shows that he is well within the tradition of scientific realism, according to which scientific theories are true descriptions of real entities that are the underlying structures of the world and that reveal its essential physical features. If the word 'true' is meant to be used in the sense of literally true, that is if the categories and statements of the true theory are thought to be in a detailed isomorphism with the world, then this position is usually called naive realism.

Naive realism faces an insurmountable difficulty, which is the underdetermination of theoretical ontology by data. As Duhem (1906) pointed out, a multiplicity of theoretical models may all fit a given set of data well, yet presuppose different ontologies (fundamental entities and properties). For example, in the context of quantum mechanics, who is able to tell which is the literally true description of the quantum world: the canonical quantization scheme in which the probability of an electron's presence is ascribed to its wave function that, according to Born (1926b), is a "ghost field" that carries no energy and momentum, and thus has no ontological status; or Feynman's path-integral method of quantization (Feynman, 1948a, 1949b; Feynman and Hibbs, 1965), in which the electron itself is an ontologically probabilistic entity and all the possible paths that this probabilistic electron can take possess equal possibility or reality? In the context of relativistic quantum mechanics, which theory is the literally true theory: Dirac's positron theory, in which a positron is a hole in the filled sea of negative energy electrons, or Feynman's theory (1949a), in which a positron is an electron evolving along the inverse arrow of time?[55] In the context of renormalization theory, which procedure is to be taken as literally true: the Pauli–Villars regularization scheme (in which unobservable auxiliary particles are introduced), or 't Hooft's (1971a, 1971b) method of dimensional regularization (in which the dimension of spacetime is taken to be a complex variable subject to analytic continuation)?

A much stronger argument against naive realism is provided by the history of science, which chronicles the frequent occurrence of conceptual revolutions in which one theory is replaced by another postulating a radically different ontology. This entails that the history of science is essentially discontinuous and supports the following metainduction: just as no fundamental theory of two hundred years ago is regarded as true today, similarly, no present day fundamental theory can have any chance to be regarded as true two hundred years later.[56]

Hacking's realism about entities is more sophisticated and emphasizes the existence of unobservable entities. The application of this kind of realism to mathematical physics, and in particular to renormalization theory, is questionable. Just think about the auxiliary particles in Pauli and Villars's regularization scheme and the Fadeev–Popov ghost in the quantization of gauge fields. Although Hacking put his emphasis on those entities that can be manipulated in experiments (including such entities as fractionally charged quarks), the boundary line separating manipulable and unmanipulable unobservable entities is quite fuzzy. Only indirect manipulations are possible and these involve a long chain of inference that presupposes a given theoretical framework. Thus the reference to the unobservable entities is far from clear cut. The defects of realism about entities lie in:

(i) the acceptance of Aristotle's realist ontology of fixed natural kinds, which is not justified or even justifiable;

(ii) the neglect of Duhem's underdetermination thesis: no data can unambiguously determine unobservable theoretical entities, so that which unobservable entities exist is always a philosophically unsolvable question; and

(iii) the underestimation of the peculiar features of mathematical physics. The predominant work in this discipline is done by thought processes and symbolic reasoning, and its aim is not to formulate a one-to-one isomorphism between the observable and unobservable entities in a theory and those in the world, but to establish a physical analogy by which the structural relations of the world can be grasped within the symbolic system that consists principally in mathematical formalisms. This is well illustrated by what Maxwell did with his mechanical model of the aether. With this model, a dynamic theory of electromagnetism was established and expressed in a set of equations. Just as an existential claim for Maxwell's idle-wheel particles is extremely dubious, so probably is Hacking's claim for a number of unobservable entities.[57]

Instrumentalists argue that if the truth claim for theories and the existence claim for unobservable entities are untenable, then scientific theories can only be regarded as mental constructions, or instruments for reasoning about the world. A scientific theory may well be coherent in its internal structure, be useful in organizing empirical materials, and be very valuable in terms of prediction, control, and explanation. Yet these traits have nothing to do with truth. Essentially, the instrumentalist argues, a scientific theory is merely a shorthand, via logic, for a complex expression that makes reference only to observed phenomena. The instrumentalist acknowledges the progress of science, but only in the sense of manifesting an increase in power as judged by the pragmatic criteria of prediction and control, not in the sense of convergence or better approximation to truth. Moreover, instrumentalists deny convergence to truth because they generally reject Aristotle's realist ontology of fixed natural kinds, so that there is nothing to converge to.[58]

Instrumentalists must face the following difficult questions: How can theories, as mental constructions mainly consisting of symbolic reasoning, be empirically adequate? Or in the same vein: What makes it possible for theories to achieve instrumental progress? These questions require a clarification of the relationship between symbols and reality. For a symbolic system to be empirically adequate, it must have some contact with reality. But mere contact is not enough. A symbolic system based on myths also has contact with reality. The essential difference between science and myth, according to realism about structural relations, lies in the fact that the structure of symbolic system in science is constrained by the structural relations in reality so that there is a correspondence between them.[59] Since instrumentalists see correspondence as taboo and so reject it, they are unable to answer adequately the questions raised above. For realists about structural relations, however, the answer is simple: the correspondence exists but it is not a one-to-one isomorphism. Rather, it is metaphoric in nature, as those that exist in models and in analogies. So the symbolic system cannot be taken as literally true, nor is the existence of any ingredient in the system guaranteed. Rather, the loose correspondence exists only between the structural relations in science, embodied as scientific laws and principles, and those in reality. As to the unobservable entities, they function as the carriers of the constellation of structural relations in a scientific theory. The possibility of rearranging the relations has deprived the entities of their primary ontological status. Illuminating examples can be found in the formalistic treatment of auxiliary particles and in the mathematical manipulation of the Higgs multiplet.

Realism about structural relations has an interesting application to equivalent theories. As is well known, there are often several theories with different unobservable entities that explain data equally well. Yet the equivalence is not absolute. One theory may provide new insights and new directions for development that people would not have thought of had they worked with the other theory. Feynman once referred to this phenomenon and stressed the different psychological implications of different theories. Rather, our explanation would refer to the differential surplus of structural relations of 'equivalent' theories: differing equivalent theories usually possess differing numbers of relations other than those required for explaining the available data, and these are carried by different unobservable entities. Another interesting application of this viewpoint is to the explanation of how a symbol, such as 'quark' in the 1960s (or the neutrino in the 1930s), can be transformed into a real entity, the quark (or the neutrino) as we now believe in it. The transformation is achieved through a change of meaning of the theoretical term from metaphoric to literal, which in turn resulted from the fact that a stable constellation of relations that had been found could be ascribed to the term.

4.5. *A Few More Remarks*

The history of renormalization, as any other history, has to be rewritten from time to time, due to changes of interests and circumstances. For example, in the late 1940s and 1950s, when the focus was on how to deal with ultraviolet divergences, a history of renormalization would narrate the events about when and how infinite results were derived within QFT (Heisenberg and Pauli, 1929; Oppenheimer, 1930; Waller, 1930; Dirac, 1934; and others), when and how the nature of the divergences – whether quadratic, linear, or logarithmic – was determined (Peierls, 1934; Weisskopf, 1934, 1939), how these divergences were shown to be identifiable, respectively, with mass and charge renormalization factors (Kramers, Bethe, Lewis, and Schwinger), and how effective techniques were developed to separate unambiguously infinities and to renormalize parameters of mass and charge (Schwinger, Tomonaga, and Feynman); and, finally, how these techniques were elevated to a research program to deal systematically with all kinds of divergences that were encountered in a perturbative calculation to arbitrary order (Dyson, Salam, Ward, Mills and Yang, Weinberg, and others). Källen, Landau, Gell-Mann and Low would be brought into the story when giving an account of the arguments, advanced in the early 1950s, indicating the internal instability of the renormalization scheme. In the triumphant mood of the early 1970s, when the renormalizable Weinberg–Salam electroweak theory enjoyed great empirical success, the history would be a record of events leading to this success: SSB and Higgs mechanism, dimensional regularization, path-integral quantization for non-Abelian gauge theories (Feynman, 1963; Fadeev and Popov, 1967; Boulware, 1970), the use of renormalization gauges ('t Hooft, 1971a, 1971b), etc. After that, physicists turned their attention to the ambitious task of constructing a final grand unified theory (Georgi and Glashow, 1974; Fritzsch and Minkowski, 1975). Since how to deal with the relevant degrees of freedom is an essential part of that enterprise, the various lines of researches that led to the renormalization group equations (Gell-Mann and Low; Fisher; Kadanoff; Wilson; Callan, 1970; and Symanzik, 1970) would become an important part of the history. Applications of SSB were an integral part of the various grand unified theories and with them came a realization of the hierarchical structure of the domains under investigation. It would take a great deal of guts for a historian of renormalization to ignore the evolution of ideas and techniques leading to the EFT approach. So in some sense, the history of renormalization, as any other internal history of science, cannot simply escape being somewhat Whiggish. However, it is still possible not to be Whiggish in the pejorative sense, if critical aspects and perspectives of the theory are provided and put into a proper intellectual context, and the problematics to be solved, when they emerge, are pointed out. Such a narrative can serve two functions. The first is to provide the story of the evolution leading to the present stage of knowledge, and the second consists in freeing physicists from the domination of current ideas that present only one of the possible ramifications. Such a history would show workers in the field the dangers of confining themselves to current ideas. As our own presentation has made clear that no development is linear or predetermined and an examination of a different perspective might be fruitful.

4.5.1. *Change of Outlook*

We have pointed out that the presence of divergences results from too naive a realism, which takes point models as literally true and globally applicable; and that the search for renormalizable theories to establish the global stability of QFT is too high an expectation. With the development of the renormalization group approach, especially with the discovery of the hierarchical structure of the domains investigated by QFT, and the development of the EFT approach for dealing practically and effectively with each layer of the hierarchy, the localist view of physical theories gained momentum, and the whole outlook about renormalizability has changed. Renormalizability seems to cease to be a necessary requirement for QFT construction, and nonrenormalizable theories can be legitimately employed.

4.5.2. *Scientific Experience and Foundational Problems*

Even though they can have long-term consequences on the internal developments of the subject, changes in the understanding of the foundations of a scientific theory do not normally have a great impact on the scientific practice and on the conception of the theory by the scientists in the field. This is because conflicting foundational schemes are not usually in conflict with the existing scientific practice, and the silent majority of scientists would normally rely on their experience rather than draw their inspiration from a new perspective suggested by a new understanding of foundational problems.[60] But, as we have tried to make clear, to decide on a choice among foundational schemes requires some expertise other than scientific experience – in particular, aptitude in conceptual analysis acquired mainly through the study of logic and philosophy, and of the history of philosophy and the philosophy of science – and the availability of historical insights obtained from the study of the history of science.

NOTES

* We would like to thank R. S. Cohen, B. Dreben, R. Jackiw, J. Polchinski, H. Putnam, J. Renn, D. Robinson, A. Shimony, and S. Weinberg for critically reading the manuscript and giving us their comments and suggestions. During the course of this work, we have benefited from helpful conversations with S. Adler, W. Bardeen, L. M. Brown, R. S. Cohen, S. Coleman, B. Dreben, M. Fisher, H. Georgi, S. Glashow, D. Gross, R. Jackiw, K. Johnson, L. Kadanoff, G. Lepage, E. Lieb, F. Low, D. Nelson, J. Polchinski, H. Putnam, J. Renn, H. Schnitzer, A. Shimony, S. Saunders, G. 't Hooft, M. Veltman, S. Weinberg, A. S. Wightman, F. Wilczek, K. Wilson, T. T. Wu, and Y. Yao. Our work has been supported in part by a grant from the National Science Foundation (DIR-9014412 (4-59070)).

[1] A stable theory is a valid theory that is insensitive to small perturbations that occur in its presumed domain, i.e., the domain that is entailed by the basic assumptions of the theory. A stable theory can withstand taking into consideration whatever happens in any part of its domain. The theory is assumed to be defined by a mathematical formalism – and the perturbations do not include changes of the formalism. The theory is assumed valid in a domain D, and initially could be defined on only a subdomain D′. A 'perturbation' might be the extension of D′ to a larger subdomain D″ of D. This initial definition of stability will be refined during the course of the paper. A more general discussion of stability will be given in Section 4.3.

An inconsistent theory is unstable because it ceases to be valid when contradictory statements appear. A consistent theory may still be unstable if in some of its presumed domain it is undefinable and undecidable. In the case of QFT, an unrenormalized theory may be consistent (i.e., it entails no self-contradictory statements) yet unstable because the occurrence of infinities indicates that it is not definable in the ultra-relativistic energy region, and the procedure of subtraction provides no effective way to decide unambiguously the truth value of some of its statements. In contrast with this, a nonrelativistic cutoff theory with a restricted domain is stable.

An unstable theory may be transformed into a quasi-stable one if the effects of the 'instabilities' can be absorbed into adjustable phenomenological parameters. In the case of QFT, a renormalizable theory seems to be quasi-stable one, while a nonrenormalizable theory is not. Yet the situation turns out to be much more complicated than it appears to be: rigorous investigations have shown that a renormalized perturbative theory is undefinable or even inconsistent (see Section 2), while a cutoff nonrenormalizable theory may be quasi-stable with a domain, presumed by the usual local field theory (see Section 3.5).

The concept of a stable theory has some affinity with the notion of a "closed theory" that Heisenberg advanced in the late 1920s and that he published in *Dialectica* in 1948. See 'The Notion of a "Closed Theory" in Modern Science' in Heisenberg (1974). For a historical account of the formulation of these ideas, see Chevalley (1988). The notion of a quasi-stable theory was also used in Schweber (1989). But the conceptualization there emphasizes the stability of the theory against small changes in the formalism. Thus it seems extremely difficult to make small changes in the formalism of quantum mechanics that maintain logical consistency, empirical validity, and the usual notions of causality. Weinberg (1989a, 1989b) has outlined an interesting nonlinear generalization of quantum mechanics that can be used as a guide in subjecting the basic framework of the theory to experimental tests. Weinberg's formalism permits an investigation of the validity of the linear superposition principle that is at the core of the usual linear version of the theory (see Bollinger et al., 1989). But Polchinski (1991) has shown that in Weinberg's version of quantum mechanics an isolated system could receive information via EPR correlations and allow EPR communication, and that, in a Stern–Gerlach experiment, communication between branches of the wave function would be possible.

[2] Certainly the tension between empirical success and conceptual instability is not unique to QFT. The case study presented in this paper serves a dual purpose: (i) to provide for philosophical analysis a contemporary example of the tension in one of the most mature and sophisticated of the physical sciences; and (ii) that the analysis in turn would have an impact on the understanding of QFT and the direction of its future development.

[3] Note that this statement, although quite popular, is by no means exact: in some cases, finite theories would also require renormalization. We thank L. M. Brown for bringing this point to our attention. An illuminating example is Landau's quasi-particle picture of Fermi liquids. In this picture, liquid ^{3}He can be described at low temperatures by a dilute gas of quasi-particle excitations near a spherical Fermi surface; and the quasi-particles can be viewed as 'dressed' versions of the constituents ^{3}He fermions, with altered masses and interactions. The basic idea here, as that adopted by field theorists, is to replace a complicated interacting system with a simpler one in which the effects of interactions are absorbed into redefinitions of masses and coupling constants.

[4] We shall not explore this aspect of the locality assumption, although it is one of the most profound problems concerning the foundations of quantum physics, and comprehensive investigations, from both physical and philosophical perspectives, have been intensively undertaken since Bell's classic papers (1964, 1966). For details and further references, compare Bell (1987).

[5] Frenkel cannot claim originality in this matter. The idea of point particles has a long tradition. We thank one of the referees of *Synthese* for reminding us of this point. For a brief exposition of the tradition within the context of classical physics, see Harman (1982a, 1982b).

[6] Note that the field operator we refer to in the text is that in the Heisenberg picture, which has different physical interpretation from that in the interaction picture.

[7] In addition, a (multiplicative) renormalization of the field operators is also necessary.

[8] We thank L. M. Brown for bringing point (ii) to our attention.

[9] This will be discussed in the next section.

[10] For more detailed descriptions of what exactly paved the way for the acceptance of Chew's *S*-matrix theory, see Cushing (1990) and Cao (1991).

[11] In the late 1960s and early 1970s, Veltman (1968a, 1968b, 1969a, 1969b, 1970) and 't Hooft (1971a, 1971b; 't Hooft and Veltman, 1972a, 1972b, 1972c) published a series of papers that paved the way for proving the renormalizability of the Yang–Mills theory in general, and that of the Weinberg–Salam model and quantum chromodynamics in particular, using dimensional regularization that preserved gauge and Poincaré invariances. The work of Veltman and 't Hooft was improved by Becchi et al. (1974). Their proof is convincing at the one-loop level; at the two-loop level, however, some people feel that the result is less convincing; as for the higher-loop levels, it seems to be virgin soil remaining to be cultivated – a remark made to us by T. T. Wu in a private conversation (Nov., 1989).

[12] At a meeting on renormalization theory in 1969, Dirac accepted the need for renormalization when looking for solutions to the Heisenberg equations of motion that are insensitive to the cutoff.

> It is necessary, for the calculations to be logical, that $\delta m/m$ and $\delta e/e$ shall be small, so that the terms involving them shall be small.... But we must not make [the cutoff] $g \to \infty$. Many physicists like to make $g \to \infty$ and believe that by doing so they get a relativistic theory. But such calculations cannot be made relativistic. (Dirac, 1969b, p. 2)

[13] This, however, is controversial. See Weinberg (1981) and Gell-Mann (1987).

[14] The divergence of the renormalized perturbative series shows the instability of renormalization theory in terms of its undefinability, without a bearing on its consistency. Yet the breaking of perturbation at ultra-relativistic energy, as shown by Landau and by Gell-Mann and Low, indicates its inconsistency. That is, assuming an expansion being perturbative entails its negation: at some energy region the expansion becomes nonperturbative.

[15] Wightman (1978) takes constructive quantum field theory as an offspring of axiomatic field theory, with some difference between them. While the concern of axiomatic field theory is the general theory of quantum fields, the constructive field theory starts from specific Lagrangian models and constructs solutions satisfying the requirements of the former. For early development of axiomatic field theory, see Jost (1965) and Streater and Wightman (1964); for constructive field theory, compare Velo and Wightman (1973).

[16] Notice that in Hilbert's formalist and finitist program a proof of consistency guarantees its stability.

[17] The following can be stated for other regularization schemes: the same claim as for the Pauli–Villars–Feynman regularization can be made for the (essentially equivalent) lattice cutoff scheme, but not for dimensional regularization, which is more formalistic and irrelevant to the point discussed here.

[18] Lepage (1989, p. 3) asserts without further explanation that "it now appears likely that this last step (taking the cutoff to infinity) is also a wrong step in the nonperturbative analysis of many theories, including QED".

[19] The nonrenormalizable interactions simulate the low energy evidence of the inaccessible high energy dynamics and are thus suppressed by a power of the experimental energy divided by the mass of the heavy boson.

[20] Explicit symmetry breaking, e.g., adding non-gauge-invariant mass terms to a pure Yang–Mills theory, is irrelevant to our discussion here.

[21] For a historical review and conceptual analysis of this subject, see Brown and Cao (1991).

[22] As pointed out by Steven Weinberg in a private correspondence (22 Sept. 1991), unitarity can be saved if one is willing to take a definition of the measure for the path integrals that spoils Lorentz invariance.

[23] An important illustration of such a constraint is the following: the standard model is renormalizable only when the number of quarks and leptons is the same, so that the anomalies caused by the quark sector and by the lepton sector cancel each other.

[24] There has been no proof for the existence of fixed points in the renormalization group transformations of the coupling constant space of QED. Thus its asymptotic scale invariance remains only a conjecture.

[25] Wilson (1983) in his Nobel lecture vividly described how the progress in statistical physics in the mid 1960s, especially the works of Widom and of Kadanoff, had influenced his thinking in theoretical physics.

[26] (a) Wilson further corroborated his dynamical view of the scale dimension of field operators with an analysis of the Thirring model (Wilson, 1970a) and of $\lambda\phi^4$ theory (Wilson, 1970b). (b) Since then even the dimensions of spacetime have become new degrees of freedom, at least in an instrumentalist sense. In statistical physics, the ϵ-expansion technique was introduced; in QFT, dimensional regularization. Both techniques are based on this new conception of spacetime dimensions. While realistic field-theoretical models ultimately have to be four-dimensional, and only toy models can be two-dimensional, in statistical physics, however, two-dimensional models are of great relevance in the real world.

[27] It is worth noting that there is an important difference between Wilson's concept of the asymptotic scale invariance of QFT at short distances and that of Bjorken (1969). While Bjorken's scaling hypothesis about the form factors in deep inelastic lepton-hadron scattering suggests that the strong interactions seem to turn off at very short distances, Wilson's formulation of OPE reestablishes the scale invariance only after absorbing the effects of interactions into the anomalous dimensions of the fields. This is just another way of expressing logarithmic corrections to the scale invariance of the theory. Thus, Bjorken's ideas were soon fitted into the framework of an non-Abelian gauge theory (QCD) and reexpressed as asymptotic freedom, while Wilson's idea has found its applications in other areas

[28] Weinberg (1983) also noticed that the relevance problem is sometimes more complicated than simply choosing an appropriate energy scale: it involves turning on collective degrees of freedom (e.g., hadrons) and turning off the elementary ones (e.g., quarks and gluons).

[29] The most widely known applications of the renormalization group equations are: (i) the observation of asymptotic freedom in QCD, which provides a basis for a perturbative QCD that is renormalizable (Politzer, 1973; Gross and Wilczeck, 1973a, 1973b); and (ii) the calculations of the variation of the coupling constants with energy in the strong and electroweak interactions, which lent support to the proposal of grand unified theories (Georgi et al., 1974).

[30] The Gaussian fixed point corresponds to a free massless field theory for which the field distributions are Gaussians.

[31] QED is regarded as perturbatively renormalizable only because the breakdown of perturbation theory at ultra-relativistic energy, as pointed out by Gell-Mann and Low, is ignored.

[32] That the tower of EFTs would be endless is entailed by the local operator formulation of QFT. Compare our earlier discussion on the operator field assumption in Section 2.

[33] This means that the nonrenormalizable effective theory can also absorb the effects of the extension of the subdomain within its domain, as presumed by a local field theory – and thus qualifies itself as a quasi-stable theory.

[34] Mary Hesse (1962, p. 262) asserts that the root of the divergence difficulties can be traced to the basic notion of atomicity. If the notion of atomicity refers to the point model or to local excitations and local couplings, the assertion is correct. However, if as usual, it refers to the structureless elementary constituents, then the root of the divergence difficulties is the structure assumption implicit in unrenormalized operator field theories, rather than the notion of atomicity.

[35] It is worth stressing that the point of departure of string theory is the possibility of considering the basic building blocks, or rather the substratum, of the universe not as zero-dimensional point particles, as they are in QFT, but as one-dimensional objects – either tiny closed loops or open curves with endpoints.

[36] Karl Popper argues convincingly about the implications of the emergence viewpoint for pluralism in theoretical ontology (see Popper, 1970, esp., pp. 6–9).

Abner Shimony (1993), in arguing for a general reductionist program, rejects the idea of objective emergence and suggests instead a notion of epistemic emergence as a compromise (see also Shimony, 1987).

[37] Part of the initial attraction of string theory was the fact that any consistent relativistic quantum theory of extended objects is highly constrained. Thus a theory of one-dimensional strings contains a quantum theory of supersymmetry, extra dimensions, gravity, gauge groups, and matter multiplets that agree with those found in the grand unification scheme of the standard model.

[38] The dearth of new experimental results until the new colliders go into operation is another factor.

[39] The localist view of scientific theory expounded in the next few paragraphs is principally due to Mary Hesse. See Hesse (1974, 1980) and Arbib and Hesse (1986).

[40] Some philosophers, in particular Ian Hacking (1983) and Nancy Cartwright (1983), maintain that, while phenomenological laws are possibly true, fundamental laws certainly lie. Part of the reason for their rejection of fundamental laws is that they play down the importance of explanation in science, which is largely the function of fundamental laws, and focus principally on predictive success. This, in turn, seems to have its roots in their 'naive' empiricist position of rejecting universals, which traditionally are defined in terms of an Aristotelian realist ontology of fixed natural kinds.

More 'sophisticated' empiricists, such as late Wittgenstein (1953) and Hesse (1974), while also rejecting Aristotelian universals, have developed an alternative theory of universals, which is based on Wittgenstein's concept of "family resemblance". They thus provided a framework within which a localist view of fundamental laws and scientific explanation can be developed (cf. Arbib and Hesse, 1986, esp. Chap 8).

[41] This position was first elaborated by Hesse (1965). Its starting point was Black's interaction theory of metaphors (1962), which was modified in the light of Wittgenstein's

"family resemblance". It was further developed by Arbib and Hess (1986). See also Hesse (1963).

[42] This point will be discussed further in Section 4.3.

[43] Fundamental physics in Japan during the 1950s and the 1960s was characterized by model-building rather than the study of dynamics. The strong influence of Taketani (1942) and Sakata (1947, 1950, 1956; Sakata and Hara, 1947; Sakata and Umezawa, 1950; Sakata et al., 1952) made Japanese physicists prefer so-called 'substantialistic' investigations over dynamical ones.

[44] Dirac's work on the relativistic theory of the electron started with his "playing" with the Pauli algebra; Gell-Mann's work on the quark model started with his playing with the $SU(3)$ λ-matrix. Cf. Dirac (1977) and Gell-Mann (1987).

[45] In choosing criteria for theory appraisal, when constructing theories some physicists have actually committed themselves to a program. For example, Weinberg takes logical inevitability, derivability from accepted principles, and simplicity as guiding principles in his theory construction. See Weinberg (1983).

[46] For a detailed discussion on the criteria of theory appraisal, compare among numerous references, Buchdahl (1970, 1980). A widely acclaimed recent discussion of the role of empirical adequacy in theory appraisal is that given by van Fraassen (1980).

[47] For example, during the nineteenth century British physicists preferred aether theories while continental physicists favored action-at-a-distance theories.

[48] In Reichenbach's terminology (1938, 1951), they are out of the context of justification.

[49] Thus, the well-defined boundary introduced by Reichenbach between the context of discovery and the context of justification is to be blurred again.

[50] Note that this criticism did not challenge the logical consistency of Newton's theory.

[51] This assertion is not undisputed in the literature (see Freudenthal, 1986). We thank one of the referees for bringing this point to our attention.

[52] Similar cases can be easily found in the history of twentieth-century physics. One example is the faith that physicists had in Pauli's hypothesis of the neutrino, prior to its detection in 1952 (see Reines, 1989) – albeit the faith was based on the more entrenched principle of energy conservation.

[53] Cf. note 15.

[54] Often the extrapolation is initiated by what Schnitzer (1988) has called a "crucial calculation".

[55] As was pointed out to us by one of the referees, note that these two theories are not completely equivalent. While Feynman's theory is applicable to bosons as well as fermions, Dirac's fails in the case of bosons. Yet this difference carries no weight in asserting that Feynman's theory of the positron is truer than Dirac's. We shall return to a further discussion of equivalent theories in the last paragraph of this section.

[56] This metainduction was first pointed out by Putnam (1977). As we have stated it, it refers to "fundamental" theories.

In a letter to us (22 Sept. 1991), Steven Weinberg, referring to the metainduction, commented:

> Even if this is true, as it may well be, there is no reason to suppose that it will continue to be true, and that we will never discover a permanent fundamental theory. After all, a nineteenth century explorer might say that "Just as no supposed source of the Nile two decades ago is regarded as the source of the Nile today, so no present

> day source of the Nile will be regarded as the true source two decades from now." Eventually the source of the Nile was discovered, and it stopped being possible to make such statements. It seems to me entirely possible that some variant of today's 'overly grandiose' superstring theories will be recognized permanently as a fundamental theory.

He also added: "I was saddened by your retreat from 'naive realism' at the end of the paper".

[57] The issue is: How is an unobservable entity rendered 'stable' or 'quasi-stable' enough to justify Hacking's claim as to their reality?

[58] The rejection of an Aristotelian realist ontology of fixed natural kinds has various ramifications. Instrumentalist epistemology is only one of them. In addition to this, it also prepares a way to a sophisticated realist epistemology. In the domain of ontology, the rejection also opens a door both for various idealisms and for dialectical materialism.

[59] Schlick (1918) and Russell (1927) discussed the reference of theoretical ontology to the structural characteristics of reality. The issue has recently been discussed by Maxwell (1970), Demopoulos and Friedman (1985), and Cao (1986).

[60] Something novel may be happening in high energy physics that is related to an interesting question in the sociology of science: Do social factors affect the *cognitive content and structure* of a scientific theory in the physical sciences? The transition from renormalizable theories to the EFT approach changes significantly the cognitive content and structure of QFT. The EFT approach involves nonrenormalizable interactions, and the determination of the limits of the applicability of some version of it justifies (in part) the building of multi-Tev accelerators such as the superconducting supercollider (SSC). This is certainly attractive to high energy experimenters connected with the SSC project. The SSC project is also attractive to high energy theorists because a large sum of funding is involved and this is important for the support of their research. Thus in the 1980s a mechanism of mutual reinforcement seemed to operate that certainly had some impact on the development of EFT approach. A social study of this episode would be interesting, but we have nothing more to report on the matter here.

REFERENCES

Adler, S. L.: 1969, 'Axial-Vector Vertex in Spinor Electrodynamics', *Phys. Rev.* **177**, 2426–38.

Aramaki, S.: 1989, 'Development of the Renormalization Theory in Quantum Electrodynamics. II', *Historia Scientiarum* **37**, 91–113.

Anderson, C. D.: 1932, 'The Apparent Existence of Easily Deflectable Positives', *Science* **76**, 328–29.

Appelquist, T., and J. Carazzone: 1975, 'Infrared Singularities and Massive Fields', *Phys. Rev.* **D11**, 2856–61.

Arbib, M. A., and M. B. Hesse: 1986, *The Construction of Reality*, Cambridge University Press, Cambridge.

Becchi, C., A. Rouet, and R. Stora: 1974, 'The Abelian Higgs–Kibble Model, Unitarity of the *S*-operator', *Phys. Lett.* **52B**, 344–46.

Bell, J. S.: 1964, 'On the Einstein–Podolsky–Rosen Paradox, *Physics* **1**, 195–200.

Bell, J. S.: 1966, 'On the Problem of Hidden Variables in Quantum Mechanics', *Rev. Mod. Phys.* **38**, 447–52.

Bell, J. S.: 1987, *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press, Cambridge.

Bell, J. S., and R. Jackiw: 1969, 'A PCAC Puzzle: $\pi^0 \rightarrow \gamma\gamma$ in the σ-model', *Nuovo Cim.* **60A**, 47–61.

Bethe, H. A.: 1947, 'The Electromagnetic Shift of Energy Levels', *Phys. Rev.* **72**, 339–41.

Bjorken, J. D.: 1969, 'Asymptotic Sum Rules at Infinite Momentum', *Phys. Rev.* **179**, 1547–53.

Black, M.: 1962, *Models and Metaphors*, Cornell University Press, Ithaca.

Bollinger, J. J., et al.: 1989, 'Test of the Linearity of Quantum Mechanics by rf Spectroscopy of the $^9\mathrm{Be}^+$ Ground State', *Phys. Rev. Lett.* **63**(10), 1031–37.

Bopp, F.: 1940, 'Eine lineare Theorie des Elektrons', *Ann. Phys.* **38**, 345–84.

Born, M.: 1926, 'Zur Quantenmechnik der Stossvorgange', *Z. Phys.* **37**, 863–67.

Born, M., W. Heisenberg, and P. Jordan: 1926, 'Zur Quantenmechnik. II', *Z. Phys.* **35**, 557–615.

Boulware, D.: 1970, 'Renormalizability of Massive Non-Abelian Gauge Fields: A Functional Integral Approach', *Ann. Phys.* **56**, 140–71.

Browder, F. E. (ed.): 1976, *Mathematical Developments Arising from Hilbert Problems*, Vol. 28, American Mathematical Society, Providence.

Brouwer, L. E. J.: 1913, 'Intuitionism and Formalism', *Bull. Amer. Math. Soc.* (*2*) **20**, 61–69.

Brown, L. M., and T. Y. Cao: 1991, 'Spontaneous Breakdown of Symmetry: Its Rediscovery and Integration into Quantum Field Theory', *Historical Studies in the Physical and Biological Sciences* **21**, 211–35.

Brown, L. M., M. Dresden, and L. Hoddeson (eds.): 1989, *Pions to Quarks*, Cambridge University Press, Cambridge.

Brown, L. M., and L. Hoddeson (eds.): 1983, *The Birth of Particle Physics*, Cambridge University Press, Cambridge.

Buchdahl, G.: 1970, 'History of Science and Criteria of Choice', in R. H. Stuewer (ed.), *Historical and Philosophical Perspective of Science*, University of Minnesota Press, Minneapolis, pp. 204–30.

Buchdahl, G.: 1980, 'Neo-transcendental Approaches towards Scientific Theory Appraisal', in D. H. Mellor (ed.), *Science, Belief and Behaviour*, Cambridge University Press, Cambridge, pp. 1–21.

Callan, C. G., Jr.: 1970, 'Bjorken Scale Invariance in Scalar Field Theory', *Phys. Rev.* **D2**, 1541–47.

Cao, T. Y.: 1986, 'A Conceptual History of Twentieth Century Field Theories', Ph.D. dissertation, University of Cambridge (to be published by Cambridge University Press in 1993–94 as *The Conceptual Developments of Twentieth Century Field Theories*).

Cao, T. Y.: 1991, 'The Reggeization Program 1962–1982: Attempts at Reconciling Quantum Field Theory with *S*-matrix Theory', *Archive for History of Exact Sciences* **41**, 239–83.

Cartan, E.: 1922, 'Sur une géneralisation de la notion de courbure de Riemann et les éspaces à torsion', *C. R. Ac. Sci.* (*Paris*) **174**, 593.

Cartwright, N.: 1983, *How the Laws of Physics Lie*, Oxford University Press, Oxford.

Chevalley, C.: 1988, 'Physical Reality and Closed Theories in Werner Heisenberg's Early Papers', in D. Batens and J. P. van Bendegem (eds.), *Theory Experiment. Recent Insights and New Perspectives on their Relations*, D. Reidel, Dordrecht, pp. 159–76.

Cohen, P. J.: 1963, 'The Independence of the Continuum Hypothesis', *Proc. Natl. Acad. Sci. (U.S.A.)* **50**, 1143–48.

Coleman, S.: 1986, *Secret Symmetry*, Cambridge University Press, Cambridge.

Coleman, S., and E. Weinberg: 1973, 'Radiative Corrections as the Origin of Spontaneous Symmetry Breaking', *Phys. Rev.* **D7**, 1888–1910.

Collins, J. C.: 1984, *Renormalization*, Cambridge University Press, Cambridge.

Collins, J. C., F. Wilczek, and A. Zee: 1978, 'Low-Energy Manifestations of Heavy Particles: Application to the Neutral Current', *Phys. Rev.* **D18**, 242–47.

Cushing, J. T.: 1990, *Theory Construction and Selection in Modern Physics: The S-Matrix Theory*, Cambridge University Press, Cambridge.

Darrigol, O.: 1986, 'The Origin of Quantized Matter Waves', *Historical Studies in the Physical and Biological Sciences* **21**, 197–253.

Davies, P. (ed.): 1989, *The New Physics*, Cambridge University Press, Cambridge.

Demopoulos, W., and M. Friedman: 1985, 'The Analysis of Matter', *Philosophy of Science* **52**, 621–39.

Dirac, P. A. M.: 1927a, 'The Quantum Theory of Emission and Absorption of Radiation', *Proc. Roy. Soc.* **A114**, 243–65.

Dirac, P. A. M.: 1927b, 'The Quantum Theory of Dispersion', *Proc. Roy. Soc.* **A114**, 716–28.

Dirac, P. A. M.: 1930, 'A Theory of Electrons and Protons', *Proc. Roy. Soc.* **A126**, 360–65.

Dirac, P. A. M.: 1934, 'Theorie du positron', in *Rapport du 7e Conseil Solvay de Physique, Structure et Proprietes des noyaux atomiques (22–29 Oct. 1933)*, Gauthier-Villars, Paris, pp. 203–12.

Dirac, P. A. M.: 1938, 'Classical Theory of Radiating Electrons', *Proc. Roy. Soc.* **A167**, 148–69.

Dirac, P. A. M.: 1942, 'The Physical Interpretation of Quantum Mechanics', *Proc. Roy. Soc.* **A180**, 1–40.

Dirac, P. A. M.: 1963, 'The Evolution of the Physicist's Picture of Nature', *Scientific American* **208**, 45–53.

Dirac, P. A. M.: 1969a, 'Methods in Theoretical Physics', in *Special Suppl. of IAEA Bulletin*, IAEA, Siena, pp. 21–28.

Dirac, P. A. M.: 1969b, 'Can Equations of Motion be Used', in *Coral Gables Conference on Fundamental Interactions at High Energy*, Gordon and Breach, New York, pp. 1–18.

Dirac, P. A. M.: 1973a, 'Relativity and Quantum Mechanics', in C. G. Sudarshan and Y. Neéman (eds.), *The Past Decades in Particle Theory*, Gordon and Breach, New York, pp. 741–72.

Dirac, P. A. M.: 1973b, 'Development of the Physicist's Conception of Nature', in J. Mehra (ed.), *The Physicist's Conception of Nature*, D. Reidel, Dordrecht, pp. 1–14.

Dirac, P. A. M.: 1977, 'Recollections of an Exciting Era', in C. Weiner (ed.), *History of Twentieth Century Physics*, Academic Press, New York, pp. 109–46.

Dirac, P. A. M.: 1983, 'The Origin of Quantum Field Theory', in Brown, Dresden, and Hoddeson (1983), pp. 39–55.

Duhem, P.: 1954, *The Aim and Structure of Physical Theory*, Princeton University Press, Princeton (originally published in 1906).

Dyson, F. J.: 1949a, 'The Radiation Theories of Tomonaga, Schwinger and Feynman', *Phys. Rev.* **75**, 486–502.

Dyson, F. J.: 1949b, 'The *S*-matrix in Quantum Electrodynamics', *Phys. Rev.* **75**, 1736–55.

Dyson, F. J.: 1951, 'The Renormalization Method in Quantum Electrodynamics', *Proc. Roy. Soc.* **A207**, 395–401.

Dyson, F. J.: 1952, 'Divergence of Perturbation Theory in Quantum Electrodynamics', *Phys. Rev.* **85**, 631–32.

Einstein, A.: 1936, 'Physics and Reality', *J. Franklin Inst.* **221**, 313–47.

Essam, J. W., and M. E. Fisher: 1963, 'Padé Approximant Studies of the Lattice Gas and Ising Ferromagnet below the Critical Point', *J. Chem. Phys.* **38**, 802–12.

Fadeev, L. D., and V. N. Popov: 1967, 'Feynman Diagrams for the Yang–Mills Field', *Phys. Lett.* **25B**, 29–30.

Feldman, D.: 1949, 'On Realistic Field Theories and the Polarization of the Vacuum', *Phys. Rev.* **76**, 1369–75.

Feynman, R. P.: 1948a, 'Space-time Approach to Non-relativistic Quantum Mechanics', *Rev. Mod. Phys.* **20**, 367–87.

Feynman, R. P.: 1948b, 'A Relativistic Cut-off for Classical Electrodynamics', *Phys. Rev.* **74**, 939–46.

Feynman, R. P.: 1948c, 'Relativistic Cut-off for Quantum Electrodynamics', *Phys. Rev.* **74**, 1430–38.

Feynman, R. P.: 1949a, 'The Theory of Positrons', *Phys. Rev.* **76**, 749–68.

Feynman, R. P.: 1949b, 'The Space-Time Approach to Quantum Electrodynamics', *Phys. Rev.* **76**, 769–89.

Feynman, R. P.: 1963, 'Quantum Theory of Gravity', *Acta Phys. Polonica* **24**, 697–722.

Feynman, R. P.: 1973, 'Partons', in Sudarshan and Neéman (1973), p. 775.

Feynman, R. P., and R. G. Hibbs: 1965, *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York.

Fisher, M. E.: 1964, 'Correlation Functions and the Critical Region of Simple Fluids', *J. Math. Phys.* **5**, 944–62.

Frenkel, J.: 1925, 'Zur Elektrodynamik punktfoermiger Elektronen', *Z. Phys.* **32**, 518–34.

Fritzsch, H., and P. Minkowski: 1975, 'Unified Interactions of Leptons and Hadrons', *Ann. Phys.* **93**, 193–266.

Freudenthal, G.: 1986, *Atom and Individual in the Age of Newton*, D. Reidel, Boston.

Furry, W. H., and J. R. Oppenheimer: 1934, 'On the Theory of the Electron and Positron', *Phys. Rev.* **45**, 245–62.

Gasiorowicz, S. G., P. R. Yennie, and H. Suura: 1959, 'Magnitude of Renormalization Constants', *Phys. Rev. Lett.* **2**, 513–16.

Gell-Mann, M.: 1987, 'Particle Theory from *S*-matrix to Quark', in M. G. Doncel, A. Hermann, L. Michel, and A. Pais (eds.), *Symmetries in Physics (1600–1980)*, Bellaterra, Barcelona, pp. 474–97.

Gell-Mann, M.,: 1989, 'Progress in Elementary Particle Theory, 1950–1964', in Brown, Dresden, and Hoddeson (1989), pp. 694–711.

Gell-Mann, M., and F. E. Low: 1954, 'Quantum Electrodynamics at Small Distances', *Phys. Rev.* **95**, 1300–12.

Gell-Mann, M., and Y. Neéman: 1964, *The Eightfold Way*, Benjamin, New York.

Georgi, H.: 1989a, 'Grand Unified Field Theories', in Davies (1989), pp. 425–45.

Georgi, H.: 1989b, 'Effective Quantum Field Theories', in Davies (1989), pp. 446–57.

Georgi, H., and S. L. Glashow: 1974, 'Unity of all Elementary Particles', *Phys. Rev. Lett.* **32**, 438–41.

Georgi, H., H. Quinn, and S. Weinberg: 1974, 'Hierarchy of Interactions in Unified Gauge Theories', *Phys. Rev. Lett.* **33**, 451–54.

Gödel, K.: 1931, 'Über formal unentscheibare Sätze der Principia Mathematica und verwandter Systeme. I', *Mh. Math. Phys.* **38**, 173–98.

Gödel, K.: 1938, 'The Consistency of the Axiom of Choice and of the Generalized Continuum-Hypothesis', *Proc. Natl. Acad. Sci.* (*U.S.A.*) **24**, 556–57.

Gödel, K.: 1939, 'Consistency-Proof for the Generalized Continuum-Hypothesis', *Proc. Natl. Acad. Sci.* (*U.S.A.*) **25**, 220–24.

Gödel, K.: 1940, 'The Consistency of the Continuum Hypothesis', *Ann. Math. Stud.*, No. 3.

Gödel, K.: 1947, 'What is Cantor's Continuum Problem?', *Am. Math. Mon.* **54**, 515–25.

Green, M. B., J. H. Schwarz, and E. Witten: 1987, *Superstring Theory*, Cambridge University Press, Cambridge.

Gross, D.: 1985, 'Beyond Quantum Field Theory', in J. Ambjorn, B. J. Durhuus, and J. L. Petersen (eds.), *Recent Developments in Quantum Field Theory*, North-Holland, Amsterdam, pp. 151–68.

Gross, D., and F. Wilczek: 1973a, 'Ultraviolet Behavior of Non-Abelian Gauge Theories', *Phys. Rev. Lett.* **30**, 1343–46.

Gross, D., and F. Wilczek: 1973b, 'Asymptotically Free Gauge Theories: I', *Phys. Rev.* **D8**, 3633–52.

Hacking, I.: 1983, *Representing and Interpreting*, Cambridge University Press, Cambridge.

Harman, P. M.: 1982a, *Energy, Force, and Matter. The Conceptual Development of Nineteenth-Century Physics*, Cambridge University Press, Cambridge.

Harman, P. M.: 1982b, *Metaphysics and Natural Philosophy. The Problem of Substance in Classical Physics*, Barnes and Noble, Totowa, N.J.

Hawking, S.: 1980, *Is the End in Sight for Theoretical Physics?*, Cambridge University Press, Cambridge.

Heisenberg, W.: 1934, 'Bemerkung zur Diracschen Theorie des Positrons', *Z. Phys.* **90**, 209–31.

Heisenberg, W.: 1960, 'Recent Research on the Nonlinear Spinor Theory of Elementary Particles', in *Proceedings of International Annual Conference on High Energy Physics*, University of Rochester, Rochester.

Heisenberg,W.: 1966, *Introduction to the Unified Field Theory of Elementary Particles*, Interscience Publishers, London.

Heisenberg, W.: 1974, *Across the Frontiers*, Harper and Row, New York.

Heisenberg, W., and W. Pauli: 1929, 'Zur Quantentheorie der Wellenfelder', *Z. Phys.* **56**, 1–61.

Heitler, W.: 1961, in R. Stoops (ed.), *The Quantum Theory of Fields*, Interscience Publishers, London.

Hesse, M. B.: 1962, *Forces and Fields*, Philosophical Library, New York.

Hesse, M. B.: 1963, *Models and Analogies in Science*, Sheed and Ward, London.

Hesse, M. B.: 1965, 'The Explanatory Function of Metaphor', in Y. Bar-Hillel (ed.), *Logic, Methodology and Philosophy of Science*, North-Holland, Amsterdam, pp. 249–59.

Hesse, M. B.: 1974, *The Structure of Scientific Inference*, MacMillan, London.

Hesse, M. B.: 1980, *Revolutions and Reconstructions in the Philosophy of Science*, Harvester, Brighton-Sussex.

Hilbert, D.: 1900, 'Mathematical Problems', in *Gottingen Nachr.* 253–97 (trans. in: 1902, *Bull. Amer. Math. Soc.* **8**, 437–79).

Hilbert, D.: 1918, 'Axiomatisches Denken', *Math. Ann.* **78**, 405–515.

Hilbert, D.: 1930, 'Probleme der Grundlagung der Mathematik', *Math. Ann.* **102**, 1–9.

Hilbert, D., and P. Bernays: 1939, *Grundlagen der Mathematik*, Vol. 2, Springer-Verlag, Berlin.

Hurst, C. A.: 1952, 'The Enumeration of Graphs in the Feynman–Dyson Technique', *Proc. Roy. Soc.* **A214**, 44–61.

Jackiw, R.: 1972, 'Field Investigations in Current Algebra', in S. B. Treiman, R. Jackiw, and D. J. Gross (eds.), *Lectures on Current Algebra and Its Applications*, Princeton University Press, Princeton, pp. 97–254.

Jaffe, A.: 1965, 'Divergence of Perturbation Theory for Bosons', *Commun. Math. Phys.* **1**, 127–49.

Johnson, K.: 1961, 'Solution of the Equations for the Green's Functions of a Two Dimensional Relativistic Field Theory', *Nuovo Cim.* **20**, 773–90.

Jordan, P., and O. Klein: 1927, 'Zum Mehrkörperproblem der Quantentheorie', *Z. Phys.* **45**, 751–65.

Jordan, P., and E. Wigner: 1928, 'Über das Paulische Equivalenzverbot', *Z. Phys.* **47**, 631–51.

Jost, R.: 1965, *The General Theory of Quantum Fields*, American Mathematical Society, Providence.

Kadanoff, L. P.: 1966, 'Scaling Laws for Ising Models near T_c', *Physics* **2**, 263–72.

Källen, G.: 1953, 'On the Magnitude of the Renormalization Constants in Quantum Electrodynamics', *Dan. Mat.-Fys. Medd.* **27**, 1–18.

Källen, G.: 1966, 'Review of Consistency Problems in Quantum Electrodynamics', *Acta Phys. Austr. Suppl.* **II**, 133–61.

Kamefuchi, S.: 1951, 'Note on the Direct Interaction between Spinor Fields', *Progr. Theor. Phys.* **6**, 175–81.

Klein, F.: 1872, *Vergleichende Betrachtungen über neuere geometrische Forschungen*, Deichert, Erlangen.

Klein, F.: 1918, 'Über die Differentialgesetze für die Erhaltung von Impuls und Energie in der Einsteinschen Gravitationstheorie', *Nachr. Ges. Wiss. Gott. Math. Phys. Kl.* 171–89.

Körner, S.: 1960, *The Philosophy of Mathematics*, Hutchinson University Library, London.

Kramers, H.: 1938, *Quantentheorie des Elektrons und der Strahlung*, Akad. Verlag, Leipzig (trans. D. ter Haar: 1957, North-Holland, Amsterdam).

Kreisel, G.: 1976, 'What have we learnt from Hilbert's Second Problem?', in Browder, (1976), pp. 93–130.

Kreisel, G.: 1980, 'Kurt Gödel', in *Biographical Memoirs of Fellows of the Royal Society* (*London*) **26**, 149–224.

Lamb, W. E., Jr., and R. C. Retherford: 1947, 'Fine Structure of the Hydrogen Atom by a Microwave Method', *Phys. Rev.* **72**, 241–43.

Landau, L. D.: 1955, 'On the Quantum Theory of Fields', in W. Pauli (ed.), *Niels Bohr and the Development of Physics*, Pergamon, London, pp. 52–69.

Landau, L. D., A. A. Abrikosov, and I. M. Khalatnikov: 1954a, 'The Removal of Infinities in Quantum Electrodynamics', *Doklady* **95**, 497–99.

Landau, L. D., A. A. Abrikosov, and I. M. Khalatnikov: 1954b, 'An Asymptotic Expression for the Electron Green Function in Quantum Electrodynamics', *Doklady* **95**, 773–76.

Landau, L. D., A. A. Abrikosov, and I. M. Khalatnikov: 1954c, 'An Asymptotic Expression for the Photon Green Function in Quantum Electrodynamics', *Doklady* **95**, 1117–20.

Landau, L. D., A. A. Abrikosov, and I. M. Khalatnikov: 1954d, 'The Electron Mass in Quantum Electrodynamics', *Doklady* **96**, 261–63.

Landau, L. D., A. A. Abrikosov, and I. M. Khalatnikov: 1956, 'On the Quantum Theory of Fields', *Nuovo Cimento Suppl.* **3**, 80–104.

Landau, L. D., and I. Pomeranchuck: 1955, 'On Point Interactions in Quantum Electrodynamics', *Doklady* **102**, 489–91.

Lee, B. W., and J. Zinn-Justin: 1972, 'Spontaneously Broken Gauge Symmetries', *Phys. Rev.* **D5**, 3121–60.

Lepage, G. P.: 1989, 'What is Renormalization?', preprint, CLNS, 89–970, Newman Lab. of Nuclear Studies, Cornell University.

Lewis, H. W.: 1948, 'On the Reactive Terms in Quantum Electrodynamics', *Phys. Rev.* **73**, 173–76.

Lie, S., and F. Engel: 1893, *Theorie der Transformationsgruppen*, Vol. 3, B. G. Teubner, Leipzig.

Lorentz, H. A.: 1904a, 'Maxwells elektromagnetische Theorie', in *Encyc. Mat. Wiss.* **V**(2), 63–144.

Lorentz, H. A.: 1904b, 'Weiterbildung der Maxwellschen Theorie: Elektronentheorie', in *Encyc. Mat. Wiss.* **V**(2), 145–280.

Mack, G.: 1968, 'Partially Conserved Dilatation Current', *Nucl. Phys.* **B5**, 499–507.

Maxwell, G.: 1970, 'Structural Realism and the Meaning of Theoretical Terms', in M. Radewer and S. Winokur (eds.), *Analyses of Theories and Methods of Physics and Psychology*, Vol. 8, Minnesota Studies in the Philosophy of Science, University of Minnesota Press, Minneapolis, pp. 181–92.

Mehra, J. (ed.): 1973, *The Physicist's Conception of Nature*, D. Reidel, Dordrecht.

Mills, R. L., and C. N. Yang: 1966, 'Treatment of Overlapping Divergences in the Photon Self-Energy Function', *Progr. Theor. Phys. Suppl.* **37–38**, 507–11.

Milne, E. A.: 1929, *The Aim of Mathematical Physics*, Oxford University Press, Oxford.

Minkowski, H.: 1908, 'Die Grundgleichungen für die elektromagnetischen Vorguge in bewegten Körper', *Gott. Nachr.* 53–111.

Minkowski, H.: 1909, 'Raum und Zeit', *Phyz. Ztschr.* **10**, 104–11.

Nafe, J. E., E. B. Nelson, and I. I. Rabi: 1947, 'The Hyperfine Structure of Atomic Hydrogen and Deuterium', *Phys. Rev.* **71**, 914–15.

Oppenheimer, J. R.: 1930, 'Note on the Interaction of Field and Matter', *Phys. Rev.* **35**, 461–77.

Pais, A.: 1945, 'On the Theory of the Electron and of the Nucleon', *Phys. Rev.* **68**, 227–28.

Pauli, W., and F. Villars: 1949, 'On the Invariant Regularization in Relativistic Quantum Theory', *Rev. Mod. Phys.* **21**, 434–44.

Pauli, W., and V. Weisskopf: 1934, 'Über die Quantisierung der skalaren relativistischen Wellengleichung'', *Helv. Phys. Acta* **7**, 709–31.

Peierls, R.: 1934, 'The Vacuum in Dirac's Theory of the Positive Electron', *Proc. Roy. Soc.* **A146**, 420–41.

Peterman, A.: 1953, 'Divergence of Perturbation Expression', *Phys. Rev.* **89**, 1160–61.

Peterman, A.: 1953, 'Renormalisation dans les séries divergentes', *Helv. Phys. Acta* **26**, 291–99.

Peterman, A., and E. C. G. Stueckelberg: 1951, 'Restriction of Possible Interactions in Quantum Electrodynamics', *Phys. Rev.* **82**, 548–49.

Pfeuti, P., and G. Toulouse: 1977, *Introduction to the Renormalization Group and Critical Phenomena*, Wiley, New York.

Poincaré, H.: 1906, 'Sur la dynamique de l'électron', *Read. Circ. Mat. Palermo* **21**, 129–75.

Polchinski, J.: 1984, 'Renormalization and Effective Lagrangians', *Nucl. Phys.* **B231**, 269–95.

Polchinski, J.: 1991: 'Weinberg's Nonlinear Quantum Mechanics and the Einstein–Podolsky–Rosen Paradox', *Phys. Rev. Lett.* **66**(4), 397–401.

Politzer, H.: 1973, 'Reliable Perturbative Results for Strong Interactions?', *Phys. Rev. Lett.* **30**, 1346–49.

Popper, K.: 1970, 'A Realist View of Logic, Physics, and History', in W. Yourgrau and A. D. Breck (eds.), *Physics, Logic, and History*, Plenum, New York, pp. 1–39.

Putnam, H.: 1977, *Meaning and the Moral Sciences*, Routledge and Kegan Paul, London.

Radicati, L. A.: 1984, 'Chaos and Cosmos', in I. Bars, A. Chodos and C.-H. Tze (eds.), *Particle Physics*, Plenum, New York, pp. 33–45.

Rayski, J.: 1948, 'On Simultaneous Interaction of Several Fields and the Self-energy Problem', *Acta Phys. Polonica* **9**, 129–40.

Reichenbach, H.: 1938, *Experience and Prediction*, University of Chicago Press, Chicago.

Reichenbach, H.: 1951, *The Rise of Scientific Philosophy*, University of California Press, Berkeley.

Reines, F.: 1989, 'Detection of the Neutrino', in Brown, Dresden, and Hoddeson (1989), pp. 359–66.

Rivier, D., and E. C. G. Stueckelberg: 1948, 'A Convergent Expression for the Magnetic Moment of the Neutron', *Phys. Rev.* **74**, 218.

Rohrlich, F.: 1973, 'The Electron: Development of the First Elementary Particle Theory', in Mehra (1973), pp. 331–69.

Russell, B.: 1914, *Our Knowledge of the External World as a Field for Scientific Method in Philosophy*, Allen and Unwin, London.

Russell, B.: 1927, *The Analysis of Matter*, Allen and Unwin, London.

Sakata, S.: 1947, 'The Theory of the Interaction of Elementary Particles', *Progr. Theor. Phys.* **2**, 145–47.

Sakata, S.: 1950, 'On the Direction of the Theory of Elementary Particles', *Iwanami* **II**, 100–03 (English trans.: 1971, *Progr. Theor. Phys. Suppl.* **50**, 155–58).

Sakata, S.: 1956, 'On a Composite Model for the New Particles', *Progr. Theor. Phys.* **16**, 686–88.

Sakata, S., and O. Hara: 1947, 'The Self-energy of the Electron and the Mass Difference of Nucleons', *Progr. Theor. Phys.* **2**, 30–31.

Sakata, S., and H. Umezawa: 1950, 'On the Applicability of the Method of Mixed Fields in the Theory of the Elementary Particles', *Progr. Theor. Phys.* **5**, 682–91.

Sakata, S., H. Umezawa, and S. Kamefuchi: 1952, 'On the Structure of the Interaction of the Elementary Particles', *Progr. Theor. Phys.* **7**, 377–90.

Salam, A.: 1951a, 'Overlapping Divergences and the S-matrix', *Phys. Rev.* **82**, 217–27.

Salam, A.: 1951b, 'Divergent Integrals in Renormalizable Field Theories', *Phys. Rev.* **84**, 426–31.

Salam, A.: 1968, 'Weak and Electromagnetic Interactions', in *Proceedings of Nobel Conference VIII*, Almquist and Wiskell, Stockholm, pp. 367–77.

Salam, A.: 1973, 'Progress in Renormalization Theory since 1949', in Mehra (1973), pp. 430–46.

Salam, A., and J. Strathdee: 1970, 'Quantum Gravity and Infinities in Quantum Electrodynamics', *Lett. Nuovo Cim.* **4**, 101–08.

Schlick, M.: 1918, *General Theory of Knowledge*, trans. A. E. Blumberg and H. Feigl, Springer-Verlag, New York.

Schnitzer, H. J.: 1988, 'The Crucial Calculation as a Motivating Force in Particle Physics', Joint Seminar for the History and Philosophy of Science, Harvard University, Cambridge.

Schweber, S. S.: 1985, 'A Short History of Shelter Island I', in R. Jackiw, N. N. Khuri, S. Weinberg, and E. Witten (eds.), *Shelter Island I*, MIT Press, Cambridge, pp. 301–43.

Schweber, S. S.: 1986a, 'Shelter Island, Pocono, and Oldstone: The Emergence of American Quantum Electrodynamics after World War II', *Osiris*, 2nd Ser. **2**, 265–302.

Schweber, S. S.: 1986b, 'Feynman and the Visualization of Spacetime Processes', *Rev. Mod. Physics* **58**, 449–508.

Schweber, S. S.: 1989, 'Molecular Beams Experiments, the Lamb Shift, and the Relation between Experiments and Theory', *Amer. J. Phys.* **57**, 299–308.

Schweber, S. S.: 1993, 'Changing Conceptualizations of Renormalization Theory', in L. Brown (ed.), *Renormalization Theory*, Springer-Verlag, New York.

Schweber, S. S., H. A. Bethe, and F. de Hoffmann: 1955, *Mesons and Fields*, Vol. I, Row, Peterson and Co., Evanston, IL.

Schwinger, J.: 1948a, 'On Quantum Electrodynamics and the Magnetic Moment of the Electron', *Phys. Rev.* **73**, 416–17.

Schwinger, J.: 1948b, 'Quantum Electrodynamics. I. A Covariant Formulation', *Phys. Rev.* **74**, 1439–61.

Schwinger, J.: 1951, 'On the Green's Functions of Quantized Field. I', *Proc. Natl. Acad. Sci. (U.S.A.)* **37**, 452–59.

Schwinger, J.: 1970, *Particles, Sources and Fields*, Vol. I, Addison-Wesley, Reading.

Schwinger, J.: 1973, 'A Report on Quantum Electrodynamics', in Mehra (1973), pp. 413–29.

Schwinger, J.: 1983, 'Renormalization Theory of Quantum Electrodynamics: An Individual View', in Brown and Hoddeson (1983), pp. 329–53.

Shimony, A.: 1987, 'Methodology of Synthesis: Parts and Wholes in Low Energy Physics', in K. Kargon and P. Achinstein (eds.), *Kelvin's Baltimore Lectures and Modern Theoretical Physics*, MIT Press, Cambridge, pp. 399–423.

Shimony, A.: 1992, 'Reality, Causality, and Closing the Circle', in his *Search for a Naturalistic World View*, Vol. 1, Cambridge University Press, Cambridge, pp. 21–61.

Streater, R. F., and A. S. Wightman: 1964, *PTC, Spin and Statistics, and All That*, Benjamin, Reading.

Stueckelberg, E. C. G.: 1938, 'Die Wechselwirkungskrafte in der Elektrodynamik und in der Feldtheorie der Kernkrafte', *Helv. Phys. Acta* **11**, 225–44, 299–329.

Stueckelberg, E. C. G., and A. Peterman: 1953, 'La normalisation des constantes dans la theorie des quanta', *Helv. Phys. Acta* **26**, 499–520.

Sudarshan, L. G., and Y. Neéman (eds.): 1973, *The Past Decade in Particular Theory*, Gordon and Breach, London.

Symanzik, K.: 1970, 'Small Distance Behavior in Field Theory and Power Counting', *Commun. Math. Phys.* **18**, 227–46.

Symanzik, K.: 1973, 'Infrared Singularities and Small-Distance-Behavior Analysis', *Commun. Math. Phys.* **34**, 7–36.

Symanzik, K.: 1983, 'Continuum Limit and Improved Action in Lattice Theories', *Nucl. Phys.* **B226**, 187–227.

Takabayasi, T.: 1983, 'Some Characteristic Aspects of Early Elementary Particle Theory in Japan', in Brown and Hoddeson (1983), pp. 294–303.

Taketani, M.: 1942, 'On Formation of the Newton Mechanics', *Kagaku*, August (trans.: 1971, *Progr. Theor.* **50**, 53–64).

Teller, P.: 1988, 'Three Problems of Renormalization', in H. R. Brown and R. Harré (eds.), *Philosophical Foundations of Quantum Field Theory*, Clarendon Press, Oxford, pp. 73–89.

Thirring, W.: 1953, 'On the Divergence of Perturbation Theory for Quantum Fields', *Helv. Phys. Acta* **26**, 33–52.

Thomson, J. J.: 1881, 'On the Electric and Magnetic Effects Produced by the Motion of Electrified Bodies', *Phil. Mag.* **11**, 227–49.

't Hooft, G.: 1971a, 'Renormalization of Massless Yang–Mills Fields', *Nucl. Phys.* **B33**, 173–99.

't Hooft, G.: 1971b, 'Renormalizable Lagrangians for Massive Yang–Mils Fields', *Nucl. Phys.* **B35**, 167–88.

't Hooft, G., and M. Veltman: 1972a, 'Renormalization and Regularization of Gauge Fields', *Nucl. Phys.* **B44**, 189–213.

't Hooft, G., and M. Veltman: 1972b, 'Combinatorics of Gauge Fields', *Nucl. Phys.* **B50**, 318–53.

't Hooft, G., and M. Veltman: 1972c, 'Example of Gauge Field Theory', in C. Korthals-Altes (ed.), *Renormalization of Yang–Mills Fields and Applications to Particle Physics*, CNRS, Marseille.

Tomonaga, S.: 1946, 'On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields', *Progr. Theor. Phys.* 1, 27–42.

Tomonaga, S.: 1965, 'Development of Quantum Electrodynamics', in *Noble Lectures (Physics): 1963–1970*, Elsevier, Amsterdam, pp. 126–36.

Umezawa, H. and R. Kawabe: 1949a, 'Some General Formulae Relating to Vacuum Polarization', *Progr. Theor. Phys.* **4**, 423–42.

Umezawa, H., and R. Kawabe: 1949b, 'Vacuum Polarization due to Various Charged Particles', *Progr. Theor. Phys.* **4**, 443–60.

Umezawa, H., J. Yukawa, and E. Yamada: 1948, 'The Problem of Vacuum Polarization', *Progr. Theor. Phys.* **3**, 317–18.

Van Fraassen, B.: 1980, *The Scientific Image*, Oxford University Press, New York.

Velo, G. and A. Wightman (eds.): 1973, *Constructive Quantum Field Theory*, Springer-Verlag, Berlin.

Veltman, M.: 1968a, 'Relation between the Practical Results of Current Algebra Techniques and the Originating Quark Model', Copenhagen Lectures, July 1968 (repr. in: R. Akhoury, B. De Wit, P. van Nieuwenhuizen, and M. Veltman (eds.): 1992, *Gauge Theory – Past and Future*, World Scientific, Singapore).

Veltman, M.: 1968b, 'Perturbation Theory of Massive Yang–Mills Fields', *Nucl. Phys.* **B7**, 637–50.

Veltman, M.: 1969a, *Proc. Topical Conf. on Weak Interactions, CERN Yellow Report*, 69-7, CERN, Geneva.

Veltman, M. (with J. Reiff): 1969b, 'Massive Yang–Mills Fields', *Nucl. Phys.* **B13**, 545–64.

Veltman, M.: 1970, 'Generalized Ward Identities and Yang–Mills Fields', *Nucl. Phys.* **B21**, 288–302.

Veltman, M.: 1977, 'Large Higgs Mass and μ–e Universality', *Phys. Lett.* **70B**, 253–54.

Waller, I.: 1930, 'Bemerküngen über die Rolle der Eigenenergie des Elektrons in der Quantentheorie der Strahlung', *Z. Phys.* **62**, 673–76.

Ward, J. C.: 1950, 'An Identity in Quantum Electrodynamics', *Phys. Rev.* **78**, 182.

Ward, J. C.: 1951, 'On the Renormalization of Quantum Electrodynamics', *Proc. Phys. Soc. (London)* **A64**, 54–56.

Weinberg, S.: 1960, 'High Energy Behavior in Quantum Field Theory', **118**, 838–49.

Weinberg, S.: 1967, 'A Model of Leptons', *Phys. Rev. Lett.* **19**, 1264–66.

Weinberg, S.: 1978, 'Critical Phenomena for Field Theorists', in A. Zichichi (eds.), *Understanding the Fundamental Constituents of Matter*, Plenum, New York, pp. 1–52.

Weinberg, S.: 1979, 'Phenomenological Lagrangians', *Physica* **96A**, 327–40.

Weinberg, S.: 1980a, 'Conceptual Foundations of the Unified Theory of Weak and Electromagnetic Interactions', *Rev. Mod. Phys.* **52**, 515–23.

Weinberg, S.: 1980b, 'Effective Gauge Theorie', *Phys. Lett.* **91B**, 51–55.

Weinberg, S.: 1983, 'Why the Renormalization Group is a Good Thing', in A. H. Guth, K. Huang, and R. L. Jaffee (eds.), *Asymptotic Realms of Physics: Essays in Honor of Francis E. Low*, MIT Press, Cambridge, pp. 1–19.

Weinberg, S.: 1989a, 'Precision Tests of Quantum Mechanics', *Phys. Rev. Lett.* **62**(5), 485–88.

Weinberg, S.: 1989b, 'Testing Quantum Mechanics', *Annals of Physics* **194**, 336–86.

Weisskopf, V. F.: 1934, 'Über die Selbstenergie des Elektrons', *Z. Phys.* **89**, 27–39.

Weisskopf, V. F.: 1936, 'Über die Elektrodynamic des Vakuums auf Grund der Quantentheorie des Elektrons', *K. Danske Vidensk. Selsk., Math.-Fys. Medd.* **14**, 1–39.

Weisskopf, V. F.: 1939, 'On the Self-energy and the Electromagnetic Field of the Electron', *Phys. Rev.* **56**, 72–85.

Weisskopf, V. F.: 1983, 'Growing up with Field Theory: The Development of Quantum Electrodynamics', in Brown and Hoddeson (1983), pp. 56–81.

Wentzel, G.: 1943, *Einfuhrung in die Quantentheorie der Wellenfelfer*, F. Deuticke, Wein.

Weyl, H.: 1918a, *Raum. Zeit. Materie*, J. Springer, Berlin.

Weyl, H.: 1918b, 'Gravitation und Elektrizität', *Sitzber. Preuss. Akad. Wiss.* 465.

Weyl, H.: 1929, 'Elektron und Gravitation', *Z. Phys.* **56**, 330–52.

Whitehead, A. N., and B. Russell: 1910–13, *Principia Mathematica*, 3 vols., Cambridge University Press, Cambridge.

Widom, B.: 1965a, 'Surface Tension and Molecular Correlations near the Critical Point', *J. Chem. Phys.* **43**, 3892–97.

Widom, B.: 1965b, 'Equation of State in the Neighborhood of the Critical Point', *J. Chem. Phys.* **43**, 3898–3905.

Wightman, A. S.: 1976, 'Hilbert's Sixth Problem: Mathematical Treatment of the Axioms of Physics', in Browder (1976), pp. 147–240.

Wightman, A. S.: 1978, 'Field Theory, Axiomatic', in *Encyclopedia of Physics*, McGraw-Hill, New York, pp. 318–21.

Wightman, A. S.: 1986, Some Lessons of Renormalization Theory', in J. de Boer, E. Dal, and D. Ulfbeck (eds.), *The Lesson of Quantum Theory*, Elsevier Science B.V., Amsterdam, pp. 201–25.

Wightman, A. S.: 1989, 'The General Theory of Quantized Fields in the 1950s', in Brown, Dresden, and Hoddeson (1989), pp. 608–29.

Wigner, E.: 1931, *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren*, F. Vieweg, Braunschweig.

Wigner, E.: 1937, 'On the Consequences of the Symmetry of the Nuclear Hamiltonian on the Spectroscopy of Nuclei', *Phys. Rev.* **51**, 106–19.

Wigner, E.: 1960, 'The Unreasonable Effectiveness of Mathematics in the Natural Sciences', *Com. Pure Appl. Math.* **13**, 1–14.

Wilson, K. G.: 1965, 'Model Hamiltonians for Local Quantum Field Theory', *Phys. Rev.* **140**, 445–57.

Wilson, K. G.: 1969, 'Non-Lagrangian Models of Current Algebra', *Phys. Rev.* **179**, 1499–1512.

Wilson, K. G.: 1970a, 'Operator-Product Expansions and Anomalous Dimensions in the Thirring Model', *Phys. Rev.* **D2**, 1473–77.

Wilson, K. G.: 1970b, 'Anomalous Dimensions and the Breakdown of Scale Invariance in Perturbation Theory', *Phys. Rev.* **D2**, 1478–93.

Wilson, K. G.: 1971, 'Renormalization Group and Strong Interactions', *Phys. Rev.* **D3**, 1818–46.

Wilson, K. G.: 1975, 'The Renormalization Group: Critical Phenomena and the Kondo Problem', *Rev. Mod. Phys.* **47**, 773–840.

Wilson, K. G.: 1983, 'The Renormalization Group and Critical Phenomena', *Rev. Mod. Phys.* **55**, 583–600.

Wilson, K. G., and M. E. Fisher: 1972, 'Critical Exponents in 3.99 Dimensions', *Phys. Rev. Lett.* **28**, 240–43.

Wilson, K. G., and J. Kogut: 1974, 'The Renomalization Group and the ϵ Expansion', *Physics Reports* **C12**, 131–264.

Wittgenstein, L.: 1922, *Tractatus Logico-Philosophicus*, ed. C. K. Ogden, Routledge and Kegan Paul, London.
Wittgenstein, L.: 1953, *Philosophical Investigations*, ed. G. E. M. Anscombe, MacMillan, London.
Yang, C. N., and R. L. Mills: 1954a, 'Isotopic Spin Conservation and a Generalized Gauge Invariance', *Phys. Rev.* **95**, 631.
Yang, C. N., and R. L. Mills: 1954b, 'Conservation of Isotopic Spin and Isotopic Gauge Invariance', *Phys. Rev.* **96**, 191–95.

Department of Physics
Brandeis University
Waltham, MA 02245
U.S.A.

and

Department of the History of Science
Harvard University
Cambridge, MA 02138
U.S.A.