

# Mathematical developments in the rise of Yang-Mills gauge theories

Adam Koberinski

December 19, 2018

**Abstract** In this paper I detail three major mathematical developments that led to the emergence of Yang-Mills theories as the foundation for the standard model of particle physics. In less than ten years, work on renormalizability, the renormalization group, and lattice quantum field theory highlighted the utility of Yang-Mills type models of quantum field theory by connecting poorly understood candidate dynamical models to emerging experimental results. I use this historical case study to provide lessons for theory construction in physics, and touch on issues related to renormalization group realism from a more historical perspective. In particular, I highlight the fact that much of the hard work in theory construction comes when trying to understand the consequences and representational capacities of a theoretical framework.

## 1 Introduction

In the mid-1960s, particle physics was in a bit of a theoretical bind. With larger and larger accelerators finding a seemingly endless supply of new resonances, the lack of a mathematically well-defined theoretical framework left theorists constructing relatively narrow phenomenological models in an attempt to systematize the new experimental findings (cf. Brown et al. (1989)). This lack of theoretical foundation did not stop the development of more sophisticated phenomenological models—like the current algebra techniques of Gell-Mann— or the discovery of important properties of certain interactions—such as the nonconservation of parity in the weak interactions. However, in the mid-1950s quantum field theory—after a brief moment of glory in the wake of the confirmation of predictions from quantum electrodynamics (QED)—quickly fell out

---

A. Koberinski  
Department of Philosophy and Rotman Institute of Philosophy, University of Western Ontario, London, ON N6A 5BY, Canada  
E-mail: akoberin@uwo.ca

of favour as a framework for particle physics. By the mid-1960s, there was no longer any candidate replacement for quantum field theory either, with S-matrix theory having lost steam by this point as well (cf. Cushing (1990)).

Over the next ten years, quantum field theory made a substantial comeback. By 1975, the Glashow-Salam-Weinberg electroweak model and quantum chromodynamics (QCD) formed the foundation of the emerging standard model of particle physics. In the intervening period, a small group of theorists continued to work on the foundations of quantum field theory—with a focus on Yang-Mills theories in particular—and discovered that the representational capacity<sup>1</sup> of the mathematical formalism was different than originally expected, and sufficiently rich to both account for known strong and weak phenomena. New dynamical models provided concrete predictions, which were subsequently tested and largely confirmed. In this paper I will outline the major mathematical discoveries regarding Yang-Mills models of quantum field theory that led to their rapid reascendancy during this period. The majority of the work done involved constructing new mathematical techniques for exploring the space of possible models of Yang-Mills theory. This is a case where foundational analysis of the theoretical framework of a discipline led to the construction of new mathematical tools for constructing dynamical models and connecting the dynamics to experiment and phenomenology.<sup>2</sup> The major techniques necessary for the acceptance of Yang-Mills models were: a full proof of the renormalizability of massless and massive Yang-Mills models, the use of renormalization group techniques to prove asymptotic freedom, and lattice quantum field theory as a tool for numerical computations in the strong coupling regime. Analysis of theory construction in the past does more than provide a philosophical depth of understanding to the history of physics; one potential contemporary use for such analysis is to learn generalizable epistemic principles that may serve as useful heuristics for theory construction today. A general lesson from this case study is that theories do not wear their consequences on their sleeve; it often takes a fair deal of analysis—including development of new mathematical tools—in order to figure out what a theory implies. These new discoveries are often more than deductive consequences of the original equations—they add to the content of the theoretical framework.

The remainder of this paper is organized as follows. In §1.1 I will introduce and explain the terminology I will use for discussing the various components of high-energy physics (HEP) in this case study. The terminology is not common, but I think it provides a more fine-grained distinction of the “pieces” of a scientific discipline, at least one that is highly mathematized. One goal of the present work is to provide a set of distinctions fine enough to account for

<sup>1</sup> By “representational capacity” I mean the ability of a model to properly capture the relevant phenomena for its domain. In the case of QCD, for example, it is important that it exhibit asymptotic freedom and quark confinement in order to match the experimental results from deep inelastic scattering and the absence of observed free quarks, respectively. I do not mean representation in some deeper sense, implying that we should therefore be realists about the model. That, I think, requires an extra step, and is outside the scope of this paper.

<sup>2</sup> These terms will be defined in §1.1.

the details of successful (and unsuccessful) theory construction, but general enough to apply outside of the domain of post-World War II particle physics.

Following this, in §2 I outline the status of Yang-Mills models circa 1965, and the epistemic context in which they were rejected. §3 provides a condensed history of the development of renormalization proofs (§3.1), renormalization group equations (§3.2), and lattice quantum field theory (§3.3). Finally, in §4, I discuss important lessons from this case study for theory construction in physics (§4.1), the use of analogies in physics (§4.2), and current discussions in the philosophy of quantum field theory (§4.3).

## 1.1 Terminology

My analysis below will split HEP into distinct components. The goal of this division is to clarify how these parts—often considered independently in philosophical analyses of scientific practice, or at least at a more coarse-grained level—work together in this successful era of theory construction. There are many ways to divide the practice of physics, and I do not claim this division is unique. It does, however, illuminate the high degree of collaboration and interaction between what is typically called theory and experiment in this era, and is sufficiently general to serve as a starting point for analysis elsewhere in physics.

By *theoretical framework*, I mean the network of principles and general mathematical constraints that serve as the common language of a research program. Currently, the theoretical framework underlying HEP is quantum field theory. In the 1960s, however, there was distrust in quantum field theory as a general framework, and so the agreed upon theoretical framework was much more minimal. It consisted of a relativistic causal structure and conservation laws carried over from non-relativistic quantum theory. Newtonian classical mechanics is another example of a theoretical framework in physics, containing concepts such as force, inertia, mass, and so on. Within a theoretical framework, one constructs *dynamical models* to describe particular interactions. As a paradigmatic example, QED constitutes a dynamical model in HEP, as it describes the electromagnetic interaction between electrons, positrons, and photons. I use dynamical model in the way most would use the term “theory,” to disambiguate the particular models of interactions from the theoretical framework guiding and constraining their construction. I include the word “dynamical” to highlight the fact that in physics, these models are often encoded in some set of dynamical evolution equations.<sup>3</sup> Typically, quantum field theory and QED would both be described as “theories,” though the latter is but an instance of the former. Given this distinction, it

---

<sup>3</sup> This is independent of the way in which dynamical models are interpreted. Dynamical models do not require a realist or mechanistic underlying interpretation. The dynamical models in the standard model—quantum chromodynamics and the electroweak model—are still the subject of heavy interpretive controversy, and many physicists involved in its construction take a clear instrumentalist view of the standard model. Nevertheless, the standard model is a clear case of a collection of dynamical models.

may be unclear what I mean by “theory construction.” For the purposes of this analysis, theory construction is the process by which a theoretical framework is established, and a consensus collection of important dynamical models emerges within that framework. For HEP, this is equivalent to the construction of the standard model and the working out of quantum field theory as its basis.

I further divide models into dynamical and phenomenological models, for a few reasons. First, the term “model” is ambiguous. In the first sense, we can understand the term as used in model theory. Then a model is simply an interpretation of a theory. Take, as an example, the theory of general relativity. Mathematically, any model of the theory is given by a specification of a tuple  $\langle \mathcal{M}, g_{\mu\nu}, T_{\mu\nu} \rangle$  including the manifold  $\mathcal{M}$ , a pseudo-Riemannian metric tensor  $g_{\mu\nu}$ , and a stress energy tensor encoding the matter content,  $T_{\mu\nu}$ . In terms of model theory, the class of these models satisfying the Einstein equations constitutes the theory of general relativity, and any particular specification is a *model* of the theory.<sup>4</sup> Hence, an FLRW cosmological solution to the Einstein equations is a model of general relativity, though it forms the basis for the theory of cosmology. This is not usually the sense of the word “model” meant in the modeling literature in philosophy of science. This second meaning usually refers to partial constructions—with input from a the relevant theory, other auxiliary theories, and perhaps phenomenology—meant to more directly model some proper subsystem that falls under a particular theory. My terminology is distinct from these two senses, though there is overlap between my phenomenological models and partial constructions. Some model-theoretic models—like those in general relativity—would also be instances of dynamical models in my sense. However, as will become clear, dynamical models in HEP are not so rigorously or formally defined.

Experiments in high-energy physics produce data, and from these data phenomena are constructed.<sup>5</sup> Phenomena are built from experimental data and expectations shaped by dynamical models or the theoretical framework. Mathematical methods and tools are used at every step of the process, in order to generate predictions, construct phenomena, and compare the two. As I will argue below, the mutual influence between experiment, mathematical tools, and a theoretical framework was essential to the construction and acceptance of QCD and the electroweak model. First I will provide a brief “state of the discipline” for HEP in this era.

---

<sup>4</sup> I use the example of general relativity here because it fits particularly well with model theory. Quantum field theory, on the other hand, is nowhere near as clearly or rigorously defined, and specifying models of quantum field theory in this sense is extremely difficult.

<sup>5</sup> The data-phenomena distinction was first explicated by Bogen and Woodward (1988). This distinction is of some importance to my view, as phenomenological models, though closely connected to experiment, can only be used to describe experimental *phenomena*, not the data. In what follows I largely gloss over this distinction, and refer to the comparison of experimental data with phenomenological models. These statements should be understood to be shorthand for the conversion of experimental data into phenomena, followed by the comparison with phenomenological models.

## 2 Yang-Mills theories in the 1960s

Before examining the theoretical developments of the late 1960s and early 1970s, it is important to understand the epistemic situation regarding Yang-Mills theories in the mid-1960s. In this section, I outline the prevailing attitudes regarding gauge freedom (§2.1) and renormalization (§2.2), and discuss the reasons for rejecting quantum field theory as an adequate framework for HEP (§2.3).

### 2.1 Gauge freedom

Yang and Mills (1954) created a construction procedure for local field theories, in analogy with QED. The idea is that one starts with a Lagrangian density describing a set of fields obeying some global internal symmetry. In QED this is the  $U(1)$  phase symmetry,  $\psi(x) \rightarrow \exp(i\alpha)\psi(x)$ ,  $\alpha \in [0, 2\pi)$ , though Yang and Mills generalize this to a transformation under a global  $SU(2)$  isospin symmetry,  $\psi(x) \rightarrow \exp(i\alpha_a t^a)\psi(x) = S\psi(x)$ , where the  $t^a$  are generators of the  $SU(2)$  symmetry group, and  $S = \exp(i\alpha_a t^a)$ . The gauge principle is a way to elevate the global symmetry group to a local symmetry group, so that the phase parameters can vary with spatiotemporal coordinates:  $\alpha_a \rightarrow \alpha_a(x)$ .<sup>6</sup> Standard Lagrangian densities involve derivatives of the fields, and these must be suitably modified to ensure that the new local symmetry leaves the Lagrangian invariant. One accomplishes this via the introduction of a covariant derivative  $D_\mu = \partial_\mu - igB_\mu$ , such that  $S[D_\mu\psi] = D_\mu\psi'$ . This amounts to the introduction of a minimal coupling to a new vector field  $B_\mu = B_{\mu a}t^a$ , whose transformation properties are constrained to be

$$B_\mu \rightarrow S^{-1}B_\mu S - \frac{i}{g}(\partial_\mu S)S^{-1}. \quad (2.1.1)$$

In QED, this new vector field is naturally assigned to the photon, such that electromagnetic interactions occur between two charged particles and a photon. More generally, the new vector field will correspond to some force mediating boson, possibly possessing internal structure. The final ingredient to complete the new so-called “gauge theory” is to introduce a kinetic energy term for the new vector field  $B_\mu$ . This is given in analogy with electromagnetism as

$$\mathcal{L}_{kin} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}; \quad F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu - [B_\mu, B_\nu]. \quad (2.1.2)$$

Yang and Mills went through the explicit construction for a generalization of isospin invariance of the strong interaction, but the procedure is easily generalized to other internal symmetry groups. The key was generalizing the

<sup>6</sup> The term “gauge” to describe the local symmetry operations comes from Weyl (1918), who sought to expand general relativity by allowing the metric field to include local variations to the length scale.

properties of the gauge field and its kinetic energy term from the Abelian  $U(1)$  group of electromagnetism to general non-Abelian groups.

The classical form of the Yang-Mills Lagrangian does not contain an explicit mass term for the gauge boson, since this term would violate gauge invariance. However, at the time of publication, Yang and Mills were uncertain of the implications for boson mass in a fully renormalized quantized theory. Due to the difficulties of renormalization, they “[had] therefore not been able to conclude anything about the mass of the  $[B]$  quantum” (p. 195). Yang and Mills argued that mass for the gauge boson was an important concern for the viability of the Yang-Mills theory as a quantum field theory of strong interactions. “[I]t is inconsistent with present experiments to have their mass less than that of the pions, because among other reasons they would then be created abundantly at high energies and the charged ones should live long enough to be seen” (p. 195).

Rather than awaiting a solution to renormalization questions for Yang-Mills theories—which wouldn’t come about until about 1970—many began adding mass terms in for gauge bosons by hand. Glashow’s (1961) early model of electroweak unification focused on a Yang-Mills type theory with an  $SU(2) \times U(1)$  gauge group. He proposed an idea of partial Lagrangian symmetries, where all terms in the Lagrangian except mass terms obeyed the partial symmetry. Symmetry concerns were important for Glashow in order to understand partially conserved currents, such as strangeness and what he called “isobaric spin.” Gauge invariance and renormalizability weren’t brought into Glashow’s discussion. He had hoped that there would be some mechanism that would ensure renormalizability and generate mass, but this wasn’t discovered until ’t Hooft proved that massless Yang-Mills theories undergoing spontaneous symmetry breaking were renormalizable, discussed below in §3.1. Gauge invariance was explicitly violated by the mass terms, and Glashow’s argument involving partial symmetries did not demonstrate that gauge invariance would be restored upon quantization. Given the analogy with QED—where gauge freedom represented mathematical redundancy and all physical quantities had to be gauge invariant—it was hard to see how mass terms that vary with gauge could be considered physical. In the next section I will discuss the status of renormalization, and how this further influenced the rejection of gauge theories.

## 2.2 Renormalization

Renormalization was initially thought of as a means to “cure” relativistic electrodynamics of its divergences. One major problem with quantizing the electromagnetic interaction was that the formalism was plagued with divergent quantities.<sup>7</sup> Arguments were given justifying the straightforward neglect of

<sup>7</sup> Divergences were also rampant in the classical relativistic theory. Unlike for nonrelativistic models of the atom—for which quantization introduced stability—quantization did not solve the divergence problem. Quantum electromagnetism suffered from logarithmic divergences instead of the steeper linear divergences of the classical theory, but the divergences

some of these infinite quantities, and subtraction procedures could be used to remove certain divergences. For example, the energy of the ground state of the Hamiltonian operator diverges in quantum field theory, but physicists at the time argued on physical grounds that only *differences in energy* from the ground state were measurable, and so one could effectively set the ground state energy to zero without changing physical predictions. In a certain sense, this “renormalizes” the vacuum energy for the theory.<sup>8</sup> However, further divergences occur within the theory, leading to two distinct problems. First, since electrodynamics is inherently relativistic, a relativistically invariant renormalization procedure was needed. If a subtraction procedure could only be carried out under a special foliation of spacetime, it was unclear if the new renormalized theory would be Lorentz invariant in the right ways. A manifestly Lorentz invariant procedure would ensure that one did not rely on a privileged reference frame. Second, the renormalization procedure could only be effective if there were a finite—and preferably a small—number of divergences to be removed from the Lagrangian or Hamiltonian. These problems were solved in tandem, independently by Tomonaga (1946), Schwinger (1948; 1949), and Feynman (1949). Dyson (1949a; 1949b) proved that the three approaches were equivalent, and strengthened the proof of renormalizability for QED. A brief account of the solutions to these problems is outlined below.<sup>9</sup>

Tomonaga (1946) developed methods for generating manifestly Lorentz-invariant generalizations of the canonical commutation relations of nonrelativistic quantum theory, and his later work uses this formalism to formulate a relativistic theory of the photon-electron interaction. His idea was to use a hybrid of the Heisenberg and Schrödinger representations for the field operators—now known as the interaction picture—to develop four-dimensional commutation relations. In a manifestly Lorentz-invariant formulation of field theory, invariant subtraction procedures for removing the infinite quantities were therefore easier to formulate. Schwinger (1948; 1949) showed that the difficult divergent quantities in QED could be absorbed into the physical electric charge—due to vacuum polarization effects—and the electron mass—due to electron self-interaction effects. He further showed that there were no analogous divergences in the photon self-energy.

---

remained. It wasn't until the advent of QED that one had a (perturbatively) divergence free formulation of relativistic electrodynamics—quantum or classical.

<sup>8</sup> In more modern terms, renormalization is a process which occurs after a regularization procedure. Regularization is a process by which divergent quantities are replaced by finite quantities depending on some arbitrary regularization parameter. Renormalization, on the other hand, is a process in which one takes the regularized theory and determines the “physical” form of the relevant parameters in the theory—usually masses and coupling constants—in such a way that they do not depend on the value of the regularization parameter. If this can be done, then the theory is renormalizable. A straightforward removal of the ground state energy value is therefore not a renormalization in this modern sense, but earlier views regarding “renormalization methods” were closer to “removing divergences from a theory.” In this older sense of the term, subtracting the ground state energy was a renormalization of the Hamiltonian.

<sup>9</sup> For a more comprehensive account of the history of the development of QED, see Schweber (1994).

Though the divergences are still problematic, the fact that they can be limited to two sources is promising for the predictive power of QED. Schwinger supposed that some future theory would cure these divergences. In the absence of this successor theory, if the divergences are limited to a small number of physical quantities, then QED can still be successful so long as they are properly renormalized away, replacing the divergent mass and charge with those “true” empirically determined values.

Feynman (1948) then developed a physical motivation for modifying QED to include the relativistic cutoff, at least in cases involving virtual photons. In a following paper, Feynman (1949) introduced the now well-known path integral formalism for QED, and showed that a relativistic cutoff procedure would yield results equivalent to Schwinger’s Hamiltonian approach when the quantities calculated could be shown to be independent of the cutoff value (i.e., were well-defined in the limit taking the cutoff to infinity). The great advantage of Feynman’s path integral approach was that it handled collision interactions in a particularly simple way. Feynman diagrams could be used to visualize the processes, and simple rules for moving from the diagrams to scattering amplitudes for collision processes were almost algorithmic in their simplicity.<sup>10</sup> Dyson (1949a) showed that the Tomonaga, Schwinger, and Feynman formalisms are all equivalent when calculations can be carried out in all three, and introduced a new renormalization procedure for Schwinger’s formalism. A few months later, Dyson (1949b) demonstrated that the formalisms of Feynman and Schwinger express the same underlying theory, insofar as their  $S$  matrix elements agree (p. 1736).

More importantly, the paper showed that QED was renormalizable—that is, it yields finite  $S$  matrix elements to all orders. The divergences are absorbed into the physical mass and charge parameters using a relativistically invariant cutoff procedure to separate out the divergent parts of a given  $S$  matrix element, and absorb them into the physical charge or mass order-by-order. Though Dyson remarks on the utility of the  $S$  matrix for prediction, he is puzzled by the structure of QED:

The surprising feature of the  $S$  matrix theory, as outlined in this paper, is its success in avoiding difficulties. Starting from the methods of Tomonaga, Schwinger and Feynman, and using no new ideas or techniques, one arrives at an  $S$  matrix from which the well-known divergences seem to have conspired to eliminate themselves. This automatic disappearance of divergences is an empirical fact, which must be given due weight in considering the future prospects of electrodynamics. Paradoxically opposed to the finiteness of the  $S$  matrix is the second fact,

<sup>10</sup> The resulting integrals, however, are not guaranteed to be amenable to analytic solution. It is still the case that diagrams with a higher order in the coupling constant  $\alpha$  lead to integrals that are enormously difficult to compute. For example, one major empirical success of QED is the high degree of precision match between the experimentally determined anomalous magnetic moment of the electron and the value predicted as a perturbative expansion of  $\alpha$  in QED. The current state of the art calculation cannot be carried out analytically, and provides a prediction to fifth order in  $\alpha$ , or tenth order in  $e$  (Kinoshita, 2013).



that the whole theory is built upon a Hamiltonian formalism with an interaction-function which is infinite and therefore physically meaningless. (p. 1754)

In order to reconcile the seeming paradox, Dyson chose to interpret the cutoff-dependent QED as representing a physical limitation to the possible measurements in the theory. If we idealize to an observer limited in measurement precision only by the principles of relativity and quantum theory (i.e.,  $\hbar$  and  $c$  as the fundamental limiting constants), then the original Hamiltonian picture would be accurate, and all physically meaningful quantities would diverge. The cutoff-dependent renormalized theory, in contrast, represents the fact that there are other physically relevant limitations on the precision of our measurements, including atomic scales and electromagnetic couplings, since our measuring equipment is composed of matter whose chief interactions are electromagnetic. Dyson shared Schwinger's hope for a more complete theory as a successor to QED, in which the divergent Hamiltonian formalism will appear as a limiting case when infinitely precise measurements are allowed. So already we see a dissatisfaction with the cutoff-based renormalization procedure, even by those who invented it. Though a successful basis for calculation, the lack of principled basis for the cutoffs was thought to represent an incompleteness to QED.

Stueckelberg and Petermann (1953) showed that the different theories one arrives at when introducing different cutoff procedures are all related, and a theory with cutoff  $\Lambda_1$  is related to one with cutoff  $\Lambda_2$  by a group transformation. This is the origin of the term "renormalization group." Gell-Mann and Low (1954) utilize their independent formulation of the renormalization group to examine the asymptotic behaviour of QED. They start by introducing a regularization parameter  $\lambda$  in such a way that the physical electric charge is a function of the bare charge and  $\lambda$ ,  $e = e_\lambda$ . The new electric charge serves to interpolate between the long-distance physical charge and the bare charge, such that  $\lim(\lambda \rightarrow 0)e_\lambda = e$  and  $\lim(\lambda \rightarrow \infty)e_\lambda = e_\infty$ . Gell-Mann and Low find that the family of parameters  $e_\lambda$  obey a differential equation of the form

$$\lambda^2 \frac{de_\lambda^2}{d\lambda^2} = f(e_\lambda^2, m^2/\lambda^2), \quad (2.2.1)$$

where the function  $f$  has a power series expansion, and the high energy values  $\lambda \gg m$  are approximated by the function  $f(e_\lambda^2, 0)$ . Using this approximation, they show that the bare charge  $e_\infty$  is either infinite, or a root of the equation  $f(e_\infty^2, 0)$ ; in both cases this is independent of the physical charge  $e$ . This is the first time that renormalization is quantitatively connected to the behaviour of QED at high energy scales, and the indication here is that the electric charge displays asymptotic divergence.

It is worth noting that most of quantum field theory in the 1950s was largely done within the canonical Hamiltonian formalism. It turns out that the choice of formalism is very important for proving renormalizability, and Yang-Mills theories are easiest to renormalize either within the path integral formalism or directly at the level of Feynman diagrams.

By the mid-1960s, those thinking about renormalization in quantum field theory would have had in mind the subtraction procedures of Tomonaga, Schwinger, Feynman, and Dyson, and perhaps the scaling behaviour of QED as investigated by Gell-Mann and Low. The reservations regarding the physical meaning of subtraction renormalization would have also been prevalent, and this was one of the reasons for a general distrust of quantum field theories as a basis for the strong and weak interactions.

### 2.3 Rejection of local field theories

Despite the empirical success of QED, by the mid-1960s most physicists were convinced of the futility of using quantum field theory as a foundation for the rest of HEP. Though the Yang-Mills procedure provided a recipe for generating field theories involving gauge bosons and nearly arbitrary internal symmetry groups—a huge advantage given the prevalence of group theory and symmetry considerations in particle classification—it appeared to only be suitable for massless gauge bosons. Further, its status as a renormalizable theory was unknown. Renormalization methods were in their infancy in the mid-1960s, and most physicists were skeptical of renormalization even in the successful domain of QED.

Both the strong and weak nuclear interactions are short range interactions, and this means that, if they are mediated by force-carrying bosons, then these bosons must be massive.<sup>11</sup> As mentioned above, adding mass terms to a Yang-Mills Lagrangian spoils gauge invariance and adds further complications to the renormalizability question.<sup>12</sup> The developments in renormalization techniques also seemed to suggest that the electric charge in QED really was divergent at high energies. Most physicists more-or-less shared the views of Dyson, Feynman, and Schwinger, and thought of QED as a useful tool for predictions, but unsuitable as a standalone fundamental theory. The simple cutoff and subtraction renormalization methods were viewed as a pragmatic way to conceal this defect of QED.

Other groups were more vocal about the rejection of quantum field theory as an inherently incoherent framework for particle physics. Landau and his collaborators (Landau et al., 1954) in Russia were also investigating the structure of QED, and found an ultraviolet pole that has since come to be known as the Landau pole. The argument is similar to the scaling behaviour investigated by Gell-Mann and Low. They show that the coupling constant

<sup>11</sup> The heuristic argument for the connection between range of interaction and mass relies on a limit based on Heisenberg's uncertainty principle. The energy-time version of the uncertainty relation is  $\Delta E \Delta t \geq 1/2\hbar$ . The timescale on which a boson can exist is related to the rest mass as  $t \approx \hbar/(2mc^2)$ . So a particle traveling near the speed of light would have a range  $R \approx \hbar/(2mc)$ . This argument is initially due to Wick (1938), explicating the Yukawa (1935) model of nuclear forces.

<sup>12</sup> Even today, the mass gap problem in Yang-Mills theories is a topic of interest among mathematically inclined physicists. The Clay Institute has offered up \$1 million as a reward for solving the mass-gap problem as one of their seven millennium problems.

for theories like QED diverges as the cutoff momentum is taken to infinity, and they interpreted this (incorrectly, as it turns out) to be a general feature of quantum field theories. Unlike the more conservative distrust of QED expressed by the American physicists who created it, Landau et al. thought their result showed that no quantum field theory could be a candidate for a complete theory of fundamental particle interactions. They took this as reason to abandon the quantum field theory formalism altogether, and this view was influential for Chew and the development of the S-matrix program (cf. (Cushing, 1990, pp.129-30)).

The S-matrix theory emerged as a rival research program for hadronic physics in the late 1950s, and its early phenomenological successes were taken by Chew, Goldberger, and others as a further sign that quantum field theory had run its course. The S-matrix theory was inspired by the mathematical problems with quantum field theory as well as the surprising successes of dealing directly with constraints on the properties of a scattering matrix. Heisenberg (1946) started an S-matrix project in the 1940s, which was largely forgotten with the rise of QED. However, Chew, Gell-Mann, and Goldberger, initially inspired by Dyson's S-matrix treatment of QED, used the S-matrix formalism as a self-consciously phenomenological framework with which to make predictions for hadronic physics (cf. Cushing (1990)). The principles of S-matrix theory became more complex, and by the mid-1960s even this rival to quantum field theory was treated with skepticism.

So, while the tools were in place that would eventually be used as the foundation of the standard model of particle physics, their properties were poorly understood. Over the next decade, however, new mathematical and phenomenological developments would lead to the rapid reemergence of Yang-Mills theories. These will be outlined in the next section

### 3 The reemergence of Yang-Mills theories

In the late-1960s, the constituent quark model was accepted as a useful fiction for classifying new hadronic particles, and Gell-Mann had begun a program of current algebra as an extension of the spirit of the S-matrix program (cf. Cao (2010)). Weinberg (1967) published a paper on a plausible field-theoretic model of electroweak unification, but this was largely ignored, for the reasons given in the previous section. There was a general distrust of quantum field theory, with the renormalizability of Yang-Mills type models still in major doubt—especially models with massive bosons. Phenomenology regarding the weak interaction was handled with a combination of current rules and the Fermi-model. Very quickly, however, experimental discoveries and mathematical advances led to the widespread acceptance of Yang-Mills theories as the foundation for both the strong and weak interactions, and quantum field theory regained its place at the foundations of particle physics.

In this section I will outline some of these developments, highlighting the ways in which the refinement and development of new mathematical tools

led to discoveries about properties of Yang-Mills theories that fit well with the emerging experimental evidence, especially for the strong interaction. §3.1 discusses the developments in understanding renormalization, leading to a proof of such for both pure and spontaneously broken Yang-Mills theories. §3.2 outlines the development of renormalization group methods, used to analyze the asymptotic behaviour of quantum field theories underlying the strong interaction. The renormalization group methods, as emphasized by Wilson, are generic tools for handling systems undergoing behaviour for which a wide range of energy scales are not only all relevant, but highly nonseparable. Finally, §3.3 outlines the lattice field theory developments, which allow for analysis of the low-energy (and therefore strong coupling) regime of the strong interaction, and provide a plausibility argument for quark confinement.

### 3.1 Proof of renormalizability

While the majority of physicists working in America and Western Europe were focused on current algebra, a few physicists retained an interest in quantum field theories. Most notable for this paper was the work of Martinus Veltman, who worked on renormalizing Yang-Mills theories in relative isolation in the late 1960s, until his student Gerard 't Hooft joined him in the early 1970s. The work of 't Hooft ended up with a proof of the renormalizability of massless Yang-Mills theories, including ones with spontaneously broken symmetries.

#### 3.1.1 Veltman's work

By the mid-1960s, the dominant formalism<sup>13</sup> in which to do quantum field theory, or to calculate S-matrix elements, was the canonical operator formalism. Within this formalism, one begins with a manifestly Poincaré-invariant Lagrangian density and uses it to construct a Hamiltonian operator in terms of the fields and their corresponding canonical field momenta. Canonical quantization proceeds by imposing (anti)commutation relations between the fields and their conjugate momenta (fermion) boson fields,

$$[\phi(x, t), \pi(y, t)]_{\pm} = i\hbar\delta^3(x - y), [\phi(x, t), \phi(y, t)]_{\pm} = [\pi(x, t), \pi(y, t)]_{\pm} = 0, \quad (3.1.1)$$

where  $\phi$  is a field operator,  $\pi = \partial_t\phi$  is its canonical momentum, and  $[\cdot, \cdot]_{\pm}$  represents the anticommutator (plus) and commutator (minus), respectively. The canonical operator formalism has the virtue of being highly similar to non-relativistic quantum mechanics, and guarantees the unitarity of the resulting S-matrix. However, in moving from working with a Lagrangian density to a Hamiltonian, Poincaré invariance becomes highly obscure. The canonical

<sup>13</sup> If one can call any formalism for quantum field theory “dominant” at this time. Trust in the reliability of quantum field theory was at an all time low, but the formalism was still used on a heuristic basis in order to arrive at the S-matrix. The S-matrix was then thought to contain all of the physical content involved in particle physics.

formalism also turns out to be ill-suited to scattering problems, where the path integral formalism excels.

In the path integral formalism, one starts with the same Lagrangian density for a classical field theory,<sup>14</sup> and inserts this into the classical action. A partition function—analogue to a partition function in statistical mechanics—is then constructed as a functional integral over field configurations in the action,

$$Z[\phi] = \int \mathcal{D}\phi \exp \left[ \frac{i}{\hbar} \int d^4x \mathcal{L}(\phi(x), \partial_\mu \phi(x)) \right], \quad (3.1.2)$$

where the term inside the exponential is  $i/\hbar$  times the classical action, and  $\mathcal{L}$  is the classical Lagrangian density. The functional integral “quantizes” the classical action by including non-extremized field configurations, i.e., “paths” for which  $\delta S[\mathcal{L}] \neq 0$ . The classical limit corresponds to focusing on the extremized action, where effectively only this one field configuration has a measurable contribution. In the path integral formalism, Poincaré invariance remains explicit, and scattering amplitudes are easily related to the functional integral. However, unlike the canonical operator formalism, unitarity is not guaranteed.

Finally, non-Abelian gauge theories complicate both formalisms significantly. Considering the fact that Yang-Mills type theories involve non-Abelian gauge fields, these complications are highly relevant for the epistemic environment of the mid-1960s. Many Russian physicists worked on modifying the path integral formalism to account for non-Abelian gauge freedom (Faddeev and Popov, 1967). Veltman, on the other hand, had the insight to work directly with the Feynman rules for a theory. Rather than trying to prove renormalizability directly from the Lagrangian, or even from one of the two dominant formalisms, Veltman found it much easier to work directly with the diagrammatic representation of a theory (Veltman, 1997). As he later recounts, however, there is an additional step when working directly with the Feynman diagrams:

a simple canonical transformation of fields may turn a perfectly reasonable set of Feynman rules into an unrenormalizable mess. Let me emphasize: unrenormalizable. An example of that is a gauge theory in the physical (or unitary) gauge. That is an unrenormalizable theory. Even if you subtract the known (that is, known from the renormalizable version) infinities, you do not wind up with a finite theory. Green’s functions have infinities all over the place. Only when you pass to the S-matrix do these infinities go away, assuming that your regularization method is quite perfect. (Veltman, 1997, p.149)

One can arrive at many, *prima facie* distinct sets of Feynman rules from the same Lagrangian through simple canonical transformations prior to deriving

---

<sup>14</sup> For quantum field theories involving fermions, the field theory has to be somewhat artificial in that Grassmann fields are used in place of classical real-valued fields. This is to ensure the appropriate anticommutation relations upon quantization.

the rules.<sup>15</sup> And even worse, renormalizability will only be provable for a small subset—perhaps a singleton set—of Feynman rules. Most sets of Feynman rules will actually be provably nonrenormalizable! This was another epistemic hurdle that Veltman had to clear: a proof of renormalizability is an existence proof in this paradigm. One must show that there *exists* a set of Feynman rules for a theory that are renormalizable. Proofs of the nonrenormalizability of a particular set of Feynman rules, which physicists at the time thought amounted to proofs of the nonrenormalizability of the dynamical model *as a whole*, do not actually tell one anything about the renormalizability of the model as a whole. As a first step to the renormalizability of Yang-Mills type theories, Veltman recounts one of the main tricks that he employed, the Bell-Treiman transformation:

Introduce a free scalar field, not interacting with the vector bosons. Now replace the vector field with some combination of vector field and scalar field; at the same time add vertices such that the scalar field remains a free field. Surely then the physics remains the same. But the Feynman rules for the new theory were different: the propagator for the W-field was replaced by the propagator for the combination, and that combination could be chosen so as to lead to less divergent Feynman rules. The price to be paid were the new vertices, and the new particle entered as a ghost (remember that is was a free particle). That is how ghosts entered my scheme. I called the technique the Bell-Treiman transformation. Neither Bell nor Treiman was responsible. (Veltman, 1997, p.155)

I reemphasize that this was only a *first step* towards a proof of the renormalizability of Yang-Mills theories.<sup>16</sup> In effect, the strategy Veltman took towards renormalization was as follows:

1. Understand the relationship between a Lagrangian and its possible sets of Feynman rules.
2. Use canonical transformations to manipulate the degrees of freedom such that gauge-varying terms end up as free fields.

---

<sup>15</sup> Note that canonical transformations are distinct from gauge transformations. A canonical transformation is a change of field variables, leading to (anti)commutation relations involving different field operators. Though this leads to the problem of unitarily inequivalent representations in quantum field theory, fields related by canonical transformations are generally thought to represent the same physical situation.

<sup>16</sup> In the context of the above quote, Veltman was actually working on renormalizing explicitly massive Yang-Mills theory, which was ultimately a failure. What Veltman accomplished was to renormalize massive Yang-Mills theory up to one loop. This was an important feat, especially in light of the more modern view of the quantum field theories as being effective theories of matter; a theory that is renormalizable to one loop can be used to generate low-energy predictions on scattering. Fermi's theory of the weak force was one-loop renormalizable, and Veltman further showed that Yang-Mills theory with an explicit mass term was equally useful. The steps Veltman took here can also apply to the massless Yang-Mills case, or the case where mass is obtained for the vector bosons through spontaneous symmetry breaking.

3. Find the “correct” set of such Feynman rules for a given Lagrangian, with which to demonstrate renormalizability.
4. Tame divergences in groups, such that renormalizability is demonstrated in steps (i.e., prove renormalizability to one-loop, then prove general renormalizability).

These techniques were then coupled with a more powerful regularization scheme by Veltman’s student, Gerard ’t Hooft.

### 3.1.2 ’t Hooft’s work

While Veltman was working on renormalizing Yang-Mills theories, a topic he “avoided dragging students into,” his student ’t Hooft expressed interest in the field. Veltman conceded, but the compromise was that, “[f]or at least part of their thesis work, I insisted on more phenomenologically oriented work” (Veltman, 1997, p.166).

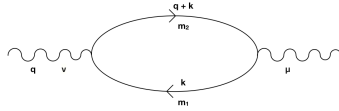
’t Hooft was initially inspired by the sigma model, constructed by Gell-Mann and Lévy (1960), which treated pions as the fundamental fields, and included a scalar sigma field whose vacuum state symmetry was spontaneously broken, giving mass to the pions. ’t Hooft suspected that spontaneous symmetry breaking may be the appropriate mass generation mechanism to ensure the renormalizability of Yang-Mills theories. This suspicion led to a particular strategy for working on the problem of renormalization: start by proving the renormalizability of massless Yang-Mills theory, and then show that the mechanism for spontaneous symmetry breaking does not spoil renormalizability. This was a departure from Veltman’s work, as Veltman was focused on explicitly massive variants of Yang-Mills theory.

The major obstacle for proving the renormalizability of massless Yang-Mills theory in general turned out to be finding a gauge invariant regularization scheme. The problem with contemporary cutoff procedures, or lattice regularization, is that gauge invariance is spoiled, and gauge invariance is required in order for the S-matrix determined from the theory’s Feynman rules to be unitary.<sup>17</sup> Further, explicitly gauge invariant regulators could be constructed, but their complexity past a one-loop correction was unwieldy, and their complexity obscured the unitarity and causality of the theory. What ’t Hooft discovered was that a new regularization trick would solve the problem in a way that allowed for easier order-by-order regularization, and which manifestly preserved the unitarity and causality of the theory: dimensional regularization.

’t Hooft started with something much like the Gell-Mann and Lévy sigma model, prior to spontaneous symmetry breaking. In effect, this is a Yang-Mills theory with an additional scalar field introduced. The first hint of the new regularization procedure came in an intermediate proof of the one-loop renormalizability of the sigma model. In order to provide appropriate counterterms for contributions to the Feynman diagrams internal to the loop, ’t Hooft moved

<sup>17</sup> As mentioned in §3.3, a gauge invariant lattice regularization procedure was eventually introduced by Wilson (1974), but was not available to ’t Hooft at the time.

to five dimensions during the regularization process. The one-loop renormalizability proof of the symmetric sigma model transferred over rather easily to the case in which the symmetry is spontaneously broken. “[T]he step remaining to be taken was a small one. As I knew from Cargèse, the actual nature of the vacuum state has little effect upon renormalization counterterms” (’t Hooft, 1997, p. 191). The transition was relatively easy because ’t Hooft realized that the regularization only needed to preserve gauge invariance of the *total set of terms* (plus counterterms) in the Lagrangian. A gauge fixing leading to spontaneous symmetry breaking amounts to the introduction of individual gauge-varying terms, though the model as a whole can remain gauge-invariant.



**Fig. 1** A single fermion loop contribution to the self-energy of the photon. From this diagram, one can easily understand the terminology of one-loop renormalization: terms like this must be properly renormalized to prove one-loop renormalizability of a dynamical model.

One can see the new regularization and renormalization processes as follows in an example from quantum electrodynamics (cf. ’t Hooft (1971a)). Consider the contribution to the photon self-energy from a single fermion loop, as in Figure 1. At a first pass, the integral associated with this diagram diverges quadratically, but can be regularized by replacing the propagator  $(m + i\gamma k)^{-1}$  with a series of terms  $\sum_j c_j (m_j + i\gamma k)^{-1}$  such that

$$c_0 = 1, \quad m_0 = m, \quad \sum_j c_j = 0, \quad \sum_j c_j m_j = 0, \quad \sum_j c_j m_j^2 = 0,$$

effectively adding a series of terms similar to the original propagator to the integral. For finite  $m_j$ , the new integral converges, and can be solved explicitly with a change of variables. Then one takes the limit of  $m_j \rightarrow \infty, j \neq 0$  (keeping the  $c_j$  constant), which is necessary in order to neglect terms in the integral of order  $q/m_j^2$ . The resulting expression, which I will refer to as  $\Pi_{\mu\nu}$ , is rather complicated, but importantly the resulting term does not satisfy gauge invariance, and the renormalized photon mass term is not zero. The gauge condition is of the following form:

$$q_\mu \Pi_{\mu\nu}(q) = 0, \quad (3.1.3)$$

where  $q_\mu$  is the photon 4-momentum.

The offending portion of the expression is a rank one polynomial in  $q^2$ , and can simply be cancelled by the introduction of a counterterm in the Lagrangian. As ’t Hooft notes, “[t]hese terms [introduced to the Lagrangian] are local and have dimension less than or equal to four, so that causality and renormalizability [respectively,] are not destroyed” (’t Hooft, 1971a, p.178). Though



a simpler renormalization scheme exists for this term in QED, 't Hooft's procedure turns out to be generalizable beyond this particular propagator and beyond QED to Yang-Mills theories; rather than imposing fully gauge-invariant constraints on the renormalization procedure, one can replace the propagator with one of revised form and add counterterms to the Lagrangian to cancel out the resulting terms that spoil gauge invariance. Arbitrary constants can then be fixed by imposing conditions of the form of equation (3.1.3), which 't Hooft calls generalized Ward identities. So long as the counterterms are local, and of dimension less than or equal to four, this procedure preserves explicit locality and renormalizability of the overall Lagrangian.

The advantage of this procedure for massless Yang-Mills theories is that the regularization is easier to define than procedures for which total gauge invariance is manifest, while the desirable properties of locality and causality are preserved. The trick in moving from QED to massless Yang-Mills is that an additional term must be added to the denominator of the propagator to regulate the Lagrangian in a gauge invariant manner. 't Hooft motivated this additional term (for one-loop) by assuming that the in-loop momenta actually have five components, and that the fifth component for all had a fixed magnitude  $M$ . This results in an effective replacement of the gauge boson and scalar particle propagators,

$$\frac{\delta_{ab}\delta_{\mu\nu}}{k^2} \rightarrow \frac{\delta_{ab}\delta_{\mu\nu}}{k^2 + M} \quad (3.1.4)$$

$$\frac{\delta_{ab}}{k^2} \rightarrow \frac{\delta_{ab}}{k^2 + M}. \quad (3.1.5)$$

Internal to the loop, the gauge boson will have a fifth polarization direction, and this is treated as a new particle with its own Feynman rules. Imposing the generalized Ward identities on a theory like this ensures renormalizability, and the new  $M$  dependence serves as an effective regulator. The case of spontaneous symmetry breaking relies on the same renormalization procedure, and 't Hooft (1971b) showed this in a follow-up paper using the example of a partial breaking of  $SU(2)$  with a scalar isospin-1 boson.<sup>18</sup>

Getting beyond one-loop renormalizability turned out to require a different approach to regularization, the hint of which is to be found in the above regularization procedure. Moving from four to five dimensions internal to one-loop was successful, but beyond one-loop renormalizability the trick was inadequate. 't Hooft and Veltman (1972) generalized the procedure into a process now known as dimensional regularization.

<sup>18</sup> In the course of proving renormalizability of Yang-Mills theories, this was a minor step. However, this second paper was hugely influential in the development of the Standard Model, as it proved at least one-loop renormalizability of the Glashow-Salam-Weinberg electroweak model. The model was therefore proven to be usable for first order predictions, which were later tested and confirmed the adequacy of the model. By this time, 't Hooft and Veltman (1972) had demonstrated full renormalizability using the dimensional regularization procedure (see below), and the electroweak model was accepted as the appropriate description of the newly unified electromagnetic and weak interactions.

The procedure suggested in [’t Hooft (1971a)] was based on the observation that the Ward identities do hold irrespective of the dimension of the space involved. By introducing a fictitious fifth dimension and a very large fifth component of momentum inside the loop suitable gauge invariant regulator diagrams could be formulated. This procedure breaks down for diagrams containing two or more closed loops because then the “fifth” component of loop momentum may be distributed over the various internal lines. It was guessed that more dimensions would have to be introduced, and thus the idea of continuation in the number of dimensions suggested itself. This is the basic idea employed in this paper. (p. 190)

’t Hooft and Veltman define an analytic continuation of S-matrix elements involving some number of loops in the complex  $n$ -plane, where positive integer values of  $n$  correspond to a spacetime dimension of that integer. The continuation is defined such that elements involving finite diagrams at  $n = 4$  agree with the conventional results. Divergences in the perturbative expansion can be shown to be poles in the complex plane at  $n = 4$ , and the generalized expressions are analytic in  $n$ . Then renormalization is just a subtraction of the poles, along with the proof of unitarity and causality of the elements for all  $n$  given by ’t Hooft and Veltman. This amounts to the claim that, at a given order in perturbation theory, the terms introduced to subtract off the pole are real and local. Unitarity uniquely determines the imaginary part of the Lagrangian from that of lower orders, so the new terms introduced at a higher order cannot contribute in unexpected ways. The requirement that the new terms are local is necessary to ensure causality.

Since the new dimensional regularization method could be applied equally well to massless Yang-Mills theory and Yang-Mills with massive gauge bosons from spontaneous symmetry breaking, ’t Hooft and Veltman proved the renormalizability of the electroweak model and what would become quantum chromodynamics. The missing ingredients for the latter—asymptotic freedom, confinement, and the presence of massless mediating bosons—were developed in parallel, and will be discussed in the next two subsections.

### 3.2 Development of the renormalization group

The work of Gell-Mann and Low (1954) on the scaling behaviour of electric charge in QED was not picked up in the particle physics community. By the mid-1960s, work on scaling was done primarily in the realm of condensed matter physics or classical statistical mechanics, where physicists’ efforts were focused on understanding critical behaviour in phase transitions.

Near the critical point of a large system, some quantity related to the system—called the order parameter—will abruptly change. At the critical point, the susceptibility of the order parameter usually diverges. The main theoretical example studied in the 1960s and 1970s was the Ising model of a ferromagnet. In the simplest version of this model, one has a cubic lattice

of spin degrees of freedom, for which the Hamiltonian includes only nearest neighbour interactions between spins (causing their direction to correlate) and the influence of an external magnetic field. At some critical temperature, the spins will spontaneously align, causing a global magnetization of the material, i.e.,  $M(T > T_C) = 0 \rightarrow M(T_C) \neq 0$ . As the system approaches the critical temperature from above, the correlation length—parameterizing the average size of “blocks” of spins that are aligned—diverges.<sup>19</sup> Kadanoff (1966) developed a technique to quantitatively analyze the spin-spin correlations for the Ising model near the critical temperature. In effect, he iterated a coarse-graining procedure by which one would take  $2 \times 2 \times 2$  blocks of spins, and treat these as contributing a single effective magnetic moment in the Hamiltonian. These blocks would then enter into the Hamiltonian with simple nearest-neighbour interactions, as in the original Ising model. The form of the Hamiltonian would remain invariant under the iterated coarse-graining, while certain physical parameters—the effective temperature and external magnetic field—might be distinct from the original model. Thus there was a set of relationships in which the temperature and external field at lattice spacing  $2L$ ,  $T_{2L}$  and  $B_{2L}$  were given in terms of  $T_L$  and  $B_L$ . Kadanoff found that, at the critical point,  $T$  and  $B$  would have fixed values, independent of the particular lattice spacing  $L$ . If the lattice spacing is allowed to change continuously, rather than in integer multiples, one can derive differential equations governing the change of  $T$  and  $B$  in terms of lattice spacing, which reach a fixed point as  $L \rightarrow \infty$ . These are formally similar to the differential equation governing the scaling behaviour of the electric charge in Gell-Mann and Low’s analysis of quantum electrodynamics. Kadanoff was one of the first to systematically treat the scaling behaviour of classical systems in the language of quantum systems. He reformulated the Onsager solution to the 2D Ising model in terms of Green’s functions. This was a key step in the eventual use of scaling methods in quantum field theory, as Green’s functions are a particularly natural tool in quantum field theory.

Kenneth Wilson, a former graduate student of Gell-Mann’s, became interested in the scaling of physical constants in quantum field theory, and began with a close analysis of the Gell-Mann and Low paper. As mentioned above in §2.2, this work used a dummy momentum index to interpolate between the long-range, physically measured electric charge  $e$ , and the short range, bare charge  $e_0$ . The renormalization group equation (2.2.1) governs the change of the electric charge  $e_\lambda$  with the change in momentum parameter  $\lambda$ , and Gell-Mann and Low show that  $e_\lambda$  increases as  $\lambda \rightarrow \infty$ , though it was unclear whether the bare charge reached a fixed point or diverged. This work inspired Wilson to consider the scaling behaviour of coupling “constants” in quantum field theory, particularly in the high-energy domain of candidate dynamical models of quantum field theory.

<sup>19</sup> This is assuming an infinitely extended ferromagnet. In general, true critical phenomena in statistical mechanics require the thermodynamic limit be taken as an idealization: the volume and number of particles both go to infinity such that the density  $N/V$  of particles remains constant.

After studying the renormalization group treatment of Gell-Mann and Low, Wilson developed a set of rules for short distance expansion for products of operators in a quantum field theory, which behave singularly in the limit of point products.<sup>20</sup> These were based on high-momentum Feynman diagrams, transformed into position space. Wilson (1965) used a similar method of analysis on the fixed source meson theory, where

I realized that the results I was getting became much clearer if I made a simplification of the fixed source model itself, in which the momentum space continuum was replaced by momentum slices. That is, I rubbed out all momenta except well separated slices, e.g.,  $1 \leq |k| \leq 2$ ,  $\Lambda \leq |k| \leq 2\Lambda$ ,  $\Lambda^2 \leq |k| \leq 2\Lambda^2$ ,  $\Lambda^n \leq |k| \leq 2\Lambda^n$ , etc. with  $\Lambda$  a large number. This model could be solved by a perturbation theory very different from the methods previously used in field theory. The energy scales for each slice were very different, namely of order  $\Lambda^n$  for the  $n$ th slice. Hence the natural procedure was to treat the Hamiltonian for the largest momentum slice as the unperturbed Hamiltonian, and the terms for all lesser slices as the perturbation. In each slice the Hamiltonian contained both a free meson energy term and an interaction term, so this new perturbation method was neither a weak coupling nor a strong coupling perturbation. (Wilson, 1982, pp.115-6)

The usual perturbation methods used in quantum field theory involve expanding in a power series about the coupling constant, while here one treats lower energy effects as perturbations to a high energy Hamiltonian. Further, Wilson generated an iterative renormalization procedure for the Hamiltonian of free meson theory. Starting with  $n$  momentum slices and using the ground state for the unperturbed  $n$ th slice Hamiltonian, the next term was an effective Hamiltonian for the remaining  $n - 1$  slices, with the coupling constant renormalized. This was the first practical use Wilson found for the renormalization group formalism. In this way, one could isolate momentum scales from the theory, solve them, and iterate to the next momentum stage.

Wilson's focus on the Hamiltonian in the fixed source meson theory transferred over to the Kadanoff picture of the Ising model, and Wilson (1971a,b) ended up reformulating and generalizing the Kadanoff picture of the Ising model near a critical point. As with much of his work on the renormalization group and critical phenomena after this point, Wilson refitted the Kadanoff picture to allow for continuous scaling, and investigated the asymptotic behaviour of differential equations relating the dependence of temperature  $T_L$  and magnetic field  $B_L$  on the scaling length  $L$ . These take the following form,

$$\frac{dT_L}{dL} = L^{-1} u(T_L, B_L^2) \quad (3.2.1)$$

$$\frac{dB_L}{dL} = L^{-1} B_L v(T_L, B_L^2), \quad (3.2.2)$$

---

<sup>20</sup> This work on operator product expansions would eventually be published as Wilson (1969), after Wilson resolved some issues with the expected behaviour of expansion coefficients in the strong coupling domain.

with the assumption that  $u$  and  $v$  were analytic at the critical temperature for a phase transition. Wilson was able to rederive the Widom-Kadanoff scaling laws for the Ising model from a more general, parallel set of assumptions, beginning with these differential equations.

In analogy with the analysis of Gell-Mann and Low, Wilson found the differential form of the renormalization group relations to be most useful. One major benefit of thinking of the renormalization group in terms of differential equations is that qualitative analyses of scaling behaviour are easily obtainable. In the case of Wilson's analysis of the Kadanoff model, the particular functions  $u$  and  $v$  that one derives are actually singular at the critical temperature. Wilson argues that the physical picture underlying the Kadanoff block spin formalism implies that the differential equations should not be singular:

In the block picture one should be able to construct  $[T_L]$  and  $[B_L]$  just by adding up interactions of individual spins within a block or across the boundary between two blocks; it is hard to see how this simple addition over a finite region can lead to singular expressions for  $[T_L]$  and  $[B_L]$ , as a function of  $[T]$  and  $[B]$ , if  $L$  is fixed. . . in the spirit of the Kadanoff approach one does not try to get specific forms for  $[u(T, B^2)]$  and  $[v(T, B^2)]$  because this would require that one take literally the idea that all spins within a block act as a unit. (Wilson, 1971a, p. 3177)<sup>21</sup>

In part II, a replacement form is found for the equations  $u$  and  $v$  to remove the singularities in the differential equation, but part I is dedicated to the qualitative features of scaling that can be determined from placing general constraints on the form of  $u$  and  $v$ . The main point is that the analytic differential equation is capable of recovering critical point singularities—most notably the Widom-Kadanoff scaling law—as asymptotic divergences at  $L = \infty$ .

Importantly for the reemergence of Yang-Mills theory, Wilson (1971c) also applied the renormalization group analysis to plausible candidates for a theory of the strong interactions. In this paper renormalization group methods are applied to the strong coupling constant, to determine how it would scale with momentum and the properties that would follow. Experiment and accompanying phenomenological models from the late 1960s indicated that scale invariance was an approximate symmetry of deep inelastic scattering—where the strong interaction was thought to play an important role—and the logarithmic corrections to scale invariance were well known (cf. (Cao, 2010, Ch.6)). Wilson demonstrated that broken scale invariance would follow if the strong interaction was described by a renormalizable theory, and if the renormalization group flow asymptotically approached a fixed point at high energy (Wilson, 1971c, Sec. III.D). The Gell-Mann and Low analysis also indicated that the electromagnetic coupling strength would increase in strength at high energies,

<sup>21</sup> Wilson uses an analogy with a simple classical mechanical system—a ball at the crest of a hill—to argue that the singularities inherent in a particular form of differential equation may be an artifact of the variables chosen to represent the equation. This is also familiar in the context of solutions to Einstein's field equations, where coordinate singularities can arise, and an appropriate transformation of coordinates must be done to remove the false singularity.

making is plausible that at some energy scale  $\Lambda$  the strength of the strong and electromagnetic forces would become equivalent. Wilson showed that, if this were true, the fixed point for strong interactions would additionally be infrared stable, allowing for an iterative “bootstrap” procedure to determine renormalized coupling constants as a function of energy scale (Sec. III.F). The importance of these results is that they provided some concrete connection between the Lagrangians describing models of the strong interaction—for which the standard perturbative procedures employed in QED would not work—and the phenomenology of strong interactions.

Conjectured in the conclusion of this paper, and confirmed by later developments, was the idea that renormalization group methods would be essential for the solution of strongly interacting relativistic fields. The benefit of the renormalization group equations is that they allow one to determine the dynamics at a particular energy scale, under the assumption that dynamics at higher energy scales have already been solved. “In order to solve the infinite number of energy scales which exceed energies of practical interest, one must iterate the renormalization-group transformation an infinite number of times, thus making asymptotic behavior of this transformation of crucial importance in solving the model” (Wilson, 1971c, p.1842).

One final scaling property of the gluon model for strong interactions—later christened QCD—was discovered by ‘t Hooft,<sup>22</sup> and later Gross and Wilczek (1973) and Politzer (1973): asymptotic freedom. The scaling of a coupling constant in massless Yang-Mills theory is parametrized by a function  $\beta(g)$  that depends on the coupling constant. Analysis of  $\beta(g)$  indicated that the coupling constant would decrease with increasing energy scales to the point where  $g \rightarrow 0$  as  $\Lambda \rightarrow \infty$ , and so a massless Yang-Mills theory is asymptotically free. For an  $SU(3)$  Yang-Mills theory, up to 16 fundamental fermion fields could be introduced without spoiling asymptotic freedom. Asymptotic freedom is important because at very high energies, the strong coupling constant would become small, and so perturbative methods could be used to extract predictions from these models. Earlier testing of scaling in the deep inelastic scattering regime (Bloom et al., 1969; Breidenbach et al., 1969)—where asymptotic freedom becomes relevant—showed that hadrons behaved as composites of point particles, and vindicated the predictions generated from the QCD Lagrangian.

### 3.3 Lattice quantum field theory

A final piece of the puzzle for connecting the QCD model with strong interaction phenomenology was required. In QCD quarks are the fundamental fermion fields, with gluons being the gauge bosons mediating the interaction between quarks. However, no free quarks or gluons had ever been detected experimentally; instead, only baryons and mesons—composed of 3 quarks or

<sup>22</sup> ‘t Hooft presented the final equation for the scaling of the beta function at a conference in 1972, but never published the results. Gross, Wilczek, and Politzer would eventually win the 2004 Nobel prize for the theoretical prediction of asymptotic freedom.

a quark-antiquark pair according to QCD, respectively—had been observed interacting via the strong force. If QCD was to be the foundational model for the dynamics of the strong interaction, it had to give rise to quark confinement in a natural way.

One problem in determining whether or not quark confinement would arise in QCD is the inverse of the asymptotic freedom property: at low energies—where baryons and mesons exist, and at which collider experiments are conducted—the strong coupling constant is large. The majority of the methods for generating dynamical predictions from quantum field theories depended on a perturbative expansion in powers of the coupling constant, and it is precisely in the most easily empirically accessible regime that these break down for QCD.

Wilson’s work on lattice QCD (1974)—inspired by, but not directly related to his work on the renormalization group analysis—provided a plausible dynamical account of quark confinement, and was convincing enough to remove this worry.

The inspiration for pursuing lattice quantum field theory began from Wilson’s earlier work on the fixed source meson model:

However, I learned from this picture of the Hamiltonian that the Hamiltonian would have to be cutoff at some large but finite value of momentum  $k$  in order to make any sense out of it, and that once it was cutoff, I basically had a lattice theory to deal with, the lattice corresponding roughly to the position space blocks for the largest momentum scale. More precisely, the sensible procedure for defining the lattice theory was to define phase space cells covering all of the cutoff momentum space, in which case there would be a single set of position space blocks, which in turn defined a position space lattice on which the field  $\phi$  would be defined. I saw from this that to understand quantum field theories I would have to understand quantum field theories on a lattice. (Wilson, 1982, p. 117)

In order to place a quantum field theory on a lattice, one must first move from a Minkowski spacetime metric to a Euclidean spatial metric.<sup>23</sup> High-momentum cutoffs in the original quantum field theory then correspond to the spatial separation of lattice sites. The biggest “trick” for developing a quantum field theory on a lattice was to make the lattice theory explicitly gauge invariant. This is important because the renormalization procedure used to make lattice quantities finite would spoil the restoration of gauge invariance afterward, so the quantities must be gauge invariant before renormalization.

In the lattice formulation of non-Abelian Yang-Mills theory coupled to fermions—of which lattice QCD is a particular model—confinement is demonstrated in the strong coupling ( $g \rightarrow \infty$ ) limit. First, one starts with a position space path integral formalism. Associated with each classical path is a contribution to the path integral which is weighted by a gauge field term, with

<sup>23</sup> This change is central to the physical disanalogies between models in quantum field theory and condensed matter physics (cf. Fraser and Koberinski (2016)). Causal and modal structures change dramatically when time is converted to a spatial dimension.

an overall weighting for all paths that includes the free gauge field action. The total weight factor is averaged over all quark paths and all gauge field values. In the strong coupling limit Wilson showed that the lattice version of the gauge field average for a closed path is exponentially suppressed by its area, meaning that the dominant contribution to the propagator is the path with least area for a given perimeter. This leads to a suppression of large area quark-antiquark loops (there is a term proportional to the gauge field average in the quark propagator) in position space; large area loops are necessary for the separation between bound quarks to be sufficient for individual detection as effectively free particles.

Though not a conclusive proof of quark confinement, Wilson provided a compelling reason to think that quarks would be confined in QCD using lattice methods. The further advantage of lattice quantum field theory is that it allows for small QCD systems to be solved using numerical methods on a computer. Even today, most results in QCD that do not depend on the use of perturbative methods in the high-energy, weak coupling limit still rely on a lattice formulation of the model.

With this last epistemic barrier removed for QCD, the path was clear for the full acceptance of the Glashow-Salam-Weinberg electroweak model and QCD as the basis for describing the fundamental forces governing the interactions in HEP.

## 4 Analysis

The rapid shift in perception of quantum field theory—from mathematically inconsistent and useless for strong interactions, to forming the basis of the standard model of particle physics—between the mid-1960s and mid-1970s has been outlined above. It is largely a story of a deepening understanding of the properties of quantum field theories (particularly of the Yang-Mills type) and their suitability as a theoretical framework for recovering particle physics phenomenology. This is a case study in developing mathematical tools to understand the properties of a theoretical framework, and then using the newly understood properties as a basis for justifying the construction of dynamical models of both the strong and (electro)weak interactions.

In this section, I will outline what I take to be the important lessons this era provides about theory construction in physics. I will start from specific epistemic and methodological breakthroughs for particle physics (§4.1 and §4.2), and finally address the relevance of looking at the process of theory construction for modern views on effective field theories (§4.3).

### 4.1 Implications for theory construction

The reemergence of Yang-Mills theory as a theoretical foundation for particle physics—forming the two dynamical models at the heart of the standard



model—highlights the importance of developing mathematical techniques for investigating the properties of a theoretical framework. In the mid-1960s theorists occupied a limited epistemic perspective on quantum field theories, particularly Yang-Mills theories. Until it was established that renormalizability was an *existence* claim for a given model of quantum field theory (i.e., one needed to find the appropriate set of Feynman rules), physicists strongly suspected that Yang-Mills theories were not renormalizable, and the possibility of a mass generation mechanism for Yang-Mills gauge bosons was not conceived. Further, the strong interaction was not well understood, and it was unclear if baryons and mesons were supposed to be the fundamental fields, or if the non-relativistic quark model could be turned into a fully relativistic field model. All of these misunderstandings are related to the limited understanding of the quantum field theoretical framework and its representational capacity.

By better understanding the representational capacity of the formal system, physicists were able to enrich their theoretical framework—by expanding the space of possible Lagrangians consistent with the core concepts of relativistic quantum field theory—and more easily construct candidate dynamical models—such as QCD and the electroweak model. Further, one could prove the compatibility of particular dynamical models with new experimental discoveries by employing the new mathematical tools. These tools were used to understand the representational capacities of the theoretical framework (quantum field theory). 't Hooft and Veltman's dimensional regularization scheme was used to show that Yang-Mills type quantum field theories were in fact renormalizable, which meant that the class of Yang-Mills models could be candidates for consistent dynamical models of strong and weak interactions. In the case of weak interactions, the Weinberg-Salam model of electroweak unification had already been worked out, so focus shifted to finding decisive empirical signatures—like weak neutral currents. For the strong force, more work was needed. As Cao (2010) has already detailed, there were multiple candidate relativistic quark-gluon models, but it was unknown how they connected to the weak binding constituent quark model—useful for group-theoretic classification of hadrons—and the relativistic current algebra.

Renormalization group analysis of the beta function for non-Abelian Yang-Mills theories showed that pure Yang-Mills theories were asymptotically free, though the addition of fermions would introduce terms in the beta function that increased with increasing energy. It turns out that the larger the Yang-Mills internal symmetry group, the more fermion generations could be accommodated without spoiling asymptotic freedom. This helped to rule out the simpler quark-gluon models, in which only a single gluon field existed (internal  $U(1)$  symmetry), and allowed QCD (internal  $SU(3)$  symmetry) to explain the success of the simple constituent quark model, in which hadrons behaved as composites of point quarks. In the deep inelastic regime, scattering results indicated that constituent quarks were effectively free, vindicating the high energy freedom of QCD. The  $SU(3)$  symmetry group matched the  $SU(3) \times SU(3)$  symmetry of current algebra as well.

The last question to be answered for QCD was whether the coupling of quarks and gluons prohibited free quarks at low energies. Wilson's lattice methods made a convincing argument for low energy quark confinement, and QCD was accepted as the consensus dynamical model of strong interactions.

Importantly, the sigma and  $U(1)$  gluon models were *not* asymptotically free, and previously proposed phenomenological models (constituent quark models, current algebra) could be connected to QCD in appropriate limiting cases. Without the new mathematical tools, these properties of candidate strong interaction models could not have been elucidated.

All three major developments outlined in §3—renormalizability, the renormalization group, and lattice quantum field theory—were tools with which to analyze the framework of quantum field theory. Renormalizability was demonstrated for large classes of quantum field theories, with the relevant models being the class of massless (or spontaneously broken) Yang-Mills theories. The knowledge that this class of models was renormalizable led to further investigations of candidate dynamical models—the electroweak model and QCD were the prime targets. The renormalization group analysis of QCD was essential to its acceptance, since asymptotic freedom made the substructure of hadrons accessible via deep inelastic scattering. Crucially, the number of fundamental fermion fields was limited in order to ensure asymptotic freedom, and this theoretical limit was consistent with the known number of fermions. The close interplay of mathematical investigation into the theoretical framework with experimental tests allowed for the emergence of consensus dynamical models within a matter of a few years.

The importance of better understanding the framework is highlighted well in Veltman's reflections on the importance of renormalizability proofs:

Personally I have always felt that the proof was much more important than the actual construction of a model, the Standard Model. I felt that, once you knew the recipe, the road to a realistic description of Nature would be a matter of time and experiment. . . The proof of renormalizability also provided detailed technical methods such as, for example, suitable regularization methods, next to indispensable for any practical application of the theory. In longer perspective, the developments in supersymmetry and supergravity have been stimulated and enhanced by the renewed respectability of renormalizable field theory. (Veltman, 1997, p.145)

Though it may seem obvious in retrospect, one lesson to keep in mind for contemporary theory construction in physics is that it takes time and innovation to discover the consequences and representation capacities of a theoretical framework. Much of the hard work in theory construction comes when trying to understand the consequences and representational capacities of a theoretical framework. In the case of particle physics, the theoretical framework of quantum field theory—mathematized in terms of Lagrangians, action functionals, canonical quantization procedures, Green's functions, scattering amplitudes, etc.—required a broader set of mathematical tools beyond perturbative ex-

pansions in coupling, and a more systematic treatment of renormalization. It turned out that the successful application of perturbative expansions and non-systematic treatments of renormalization were permissible for QED due to some peculiar features of the electromagnetic interaction: first, the dimensionless coupling constant for QED is relatively small ( $\alpha \approx 1/137$ ), making a perturbation about the coupling constant accurate after only a few orders; second, QED is a special case of an Abelian gauge theory, and did not require more sophisticated regularization and renormalization techniques. The failure of techniques that worked for the comparatively simple QED did not mean that quantum field theory in general would fail, and it took the work of the few physicists who remained interested in quantum field theory to prove this.

One important feature of quantum field theory in the mid-1960s was that there was already a useful dynamical model in existence: QED. Though it seemed to be the case that QED was not easily extended to accommodate the strong or weak interactions—for example, the renormalization procedures were not easily generalized—it provided a clear example showing that quantum field theory was at least suitable for a limited domain of particle physics. In the context of quantum field theory, we can now see why straightforward extensions of QED would have to be unsuccessful. One needed to move from Abelian gauge theory to non-Abelian gauge theory, and the tools required to handle the latter turned out to be much more complicated.

#### 4.2 The success of formal analogies with statistical mechanics

An understanding of the full representational capacity of Yang-Mills theories required the development of novel mathematical techniques to explore their renormalizability, scaling behaviour, numerical solutions, and mass generation mechanisms.<sup>24</sup> All of these methods were either developed within or originated from condensed matter physics, and were carried (back) over to particle physics due to the formal similarities between the structure of models in the two disciplines.

As mentioned in §3.2 and §3.3, much of Wilson’s work regarding the renormalization group and lattice QCD was inspired by strong formal analogies with statistical mechanical systems. Though the initial application of renormalization group equations to particle physics was through Gell-Mann and Low’s (1954) treatment of scaling in QED, condensed matter physicists (like Kadanoff) did much of the work on the renormalization group in the 1960s, in the context of simple statistical mechanical models—both quantum *and* classical. Many of Wilson’s papers on the renormalization group deal variously with particle physics and statistical physics. The three landmark papers published in 1971 deal with the applicability of renormalization group methods to the strong interaction (1971a), and to the Kadanoff scaling picture of

<sup>24</sup> Mass generation—in the form of spontaneous symmetry breaking—was not discussed in this paper. For a detailed analysis of the formal analogy between spontaneous symmetry breaking in the Higgs mechanism and in superconductivity, see Fraser and Koberinski (2016).

the Ising model (1971b;1971c). In the more systematic paper on renormalization group methods, Wilson and Kogut (1974), both statistical mechanical and quantum field theoretic systems are treated. A few sections at the end of their paper—particularly Section 10—outline the formal connection between statistical mechanics and quantum field theory. In particular, the Feynman diagrams for a  $\phi^4$  lattice quantum field theory—when converted to a Euclidean metric—are identical to the spin correlation functions for the generalized Ising model.

Wilson and Kogut (1974, Sec 12.2) present the details of renormalization group flow and its applicability to quantum field theory in the case of a four-dimensional  $\phi^4$  model. One starts with the model defined on a lattice, which is implied by the introduction of a momentum cutoff scale. First, one must regularize the model, and then introduce a nonzero coupling constant at infinite “correlation length”—the analogue of which in quantum field theory is a continuum relativistic model. The reasons that a tight formal correspondence can be set up between (classical, in this case) statistical mechanics and quantum field theory are complex, but the application of the renormalization group analysis to both is no accident. In later works, Wilson explicitly emphasizes the expected generality of renormalization group methods as applicable to numerical solutions of physical situations. Unlike the simple problems with a few degrees of freedom that are amenable to simple numerical approximation schemes,

There is a fourth class of problems which until very recently lacked *any* convincing numerical approach. This fourth class is a subclass of problems involving a large or infinite number of degrees of freedom. The special feature of this subclass is the problem of “renormalization.” Originally, renormalization was the procedure for removing the divergences of quantum electrodynamics and was applied to the Feynman graph expansion. The difficulties of renormalization prevent one from formulating even a Monte Carlo method for quantum electrodynamics. Similar difficulties show up in a number of problems scattered throughout physics (and chemistry, too). These problems include: turbulence (a problem in classical hydrodynamics), critical phenomena (statistical mechanics), dilute magnetic alloys, known as the Kondo problem (solid state physics), the molecular bond for large molecules (chemistry), in addition to all of quantum field theory.

In this paper the problem of renormalization will be shown to be the problem of *many length* or *energy scales*. (Wilson, 1975, p.171)

It is clear from this quote that Wilson, at least, saw the renormalization group as a mathematical technique for systematically treating problems where energy scales are not cleanly separable. So the treatment of renormalization, though initially developed in order to make QED a predictive model, is really a quite general phenomena in physics, and should perhaps be likened more to techniques such as Taylor expansions and Sturm-Liouville theory. These are

techniques for finding solutions to formal problems, that have little directly to do with the physical situation at hand.

So if the renormalization group methods are meant to be so generally applicable, why was the analogy with statistical mechanics important for their development beyond QED in particle physics? It's because the trick with the renormalization group is to find a way to express the problem that makes it clear how to use the techniques.

[I]t is rather difficult to formulate renormalization group methods for new problems; in fact, the renormalization group approach generally seems as hopeless as any other approach until someone succeeds in solving the problem by the renormalization group approach. Where the renormalization group approach has been successful a lot of ingenuity has been required: one cannot write a renormalization group cookbook. (Wilson, 1975, p.185).

After Wilson's success in formulating the Ising model in such a way that a renormalization group analysis could be performed, he looked for ways to transform quantum field theory problems into the same form as the classical statistical mechanical Ising model. As Fraser (2018) has emphasized, the analogy between the classical Ising model and four-dimensional  $\phi^4$  model is facilitated by transforming the  $\phi^4$  model into a Euclidean metric by Wick rotating the temporal part of the metric  $t \rightarrow -it$ , and discretizing the spacetime by introducing a minimum length scale. Then, one must establish a formal correspondence between the spin-spin correlation function  $\Gamma_{m,n}$  in the Ising model and the propagator  $D_m(n\tau)$  in the  $\phi^4$  model,

$$\Gamma_{m,n} = \zeta D_m(n\tau), \quad (4.2.1)$$

where the Ising model is defined on a lattice with spacings  $m$  and  $n$ , the propagator is defined on a lattice with spatial spacing  $m$  and temporal spacing  $n$ ,  $\tau = -it$ , and  $\zeta$  is a constant of proportionality. Given this formal identification of models in statistical mechanics, the renormalization group formulation of the Ising model can be applied straightforwardly to quantum field theory.<sup>25</sup>

To summarize, Wilson developed methods for solving models in physics for which energy scales are all highly linked, and qualitative behaviour is sensitive to the interactions spanning large ranges of energy. In order to apply these methods to non-Abelian quantum field theory—particularly Yang-Mills models—formal analogies with soluble models in statistical mechanics were essential. Wilson thus used the successes in statistical mechanics as a basis for formal analogies with quantum field theory, and found ways to apply the renormalization group analysis across both domains. In setting up the formal correspondence, putting quantum field models on a lattice was an important intermediate step as well. The importance of the inspiration from statistical

<sup>25</sup> There is a bit more work to be done to establish that the quantum field model reaches a critical surface when the continuum limit is taken, and this will vary from model to model within quantum field theory. See Wilson and Kogut (1974); Fraser (2018) for the remaining details.

mechanics was that renormalization problems had already been successfully solved there; Wilson was able to more-or-less carry over those results once he had established the identity in equation (4.2.1).

### 4.3 Renormalizability and effective field theory

Given the predominant viewpoint in modern physics that quantum field theories—and the standard model in particular—form an adequate framework only of *effective field theories*, the importance of renormalizability proofs for the acceptance of Yang-Mills theories might seem a bit odd. If we view quantum field theories as effective theories with a built in cutoff scale (Weinberg, 1979; Wallace, 2011; Williams, 2017; Fraser, 2017), then full renormalizability cannot be a necessary mathematical property of a physical model of quantum field theory, however nice it may be.<sup>26</sup> So why was the HEP community so dismissive of candidate Yang-Mills models for the strong and (electro)weak interactions until 't Hooft published a proof of their renormalizability?

The most important reason that a full proof of the renormalizability of Yang-Mills theories was essential to their acceptance is that the view of the standard model as a collection of effective field theories depends critically on the use of renormalization group methods to demonstrate that non-renormalizable terms become negligible at low energies. Weinberg (1979) outlined the utility of the use of what he called “phenomenological Lagrangians” as a tool for both understanding properties of interactions not captured within known models, and justifying the addition of nonlinear, nonrenormalizable terms. Though fully nonrenormalizable terms in a Lagrangian can only be used to make predictions at low orders—higher order contributions from such terms introduce a growing number of arbitrary parameters that must be fixed—one can study their scaling behaviour using the renormalization group methods. In this way, high energy properties—both qualitative *and* quantitative—of nonrenormalizable terms can be explored. Further, a general lesson from the renormalization group analysis of critical phenomena in statistical mechanics is that, at large distances (equivalent to low energy-momenta) many interaction terms become entirely irrelevant to the behaviour of the system in question. That is, the relative contributions from certain interaction terms “die off” at low energies,

<sup>26</sup> There is not a consensus that an effective field theory view of the standard model is the best way to interpret the utility of quantum field theoretic models. Many people working in axiomatic and/or algebraic quantum field theory, for example, aim to provide an exact model for realistic interactions, to which the standard perturbative methods of conventional quantum field theory create an asymptotic expansion (e.g., Streater and Wightman (1964); Buchholz and Verch (1995); Halvorson and Müger (2007); Feintzeig (2017)). These may still be effective theories in the sense that they have a limited domain of applicability, but they would then be candidates for a more standard philosophical interpretation. Others have criticized the limited utility of a realist interpretation of effective field theories based on the renormalization group (Fraser, 2018; Ruetsche, 2018). Though these are important philosophical issues, they are orthogonal to the discussion here. For the purposes of the main discussion, I will uncritically accept the effective field theory view, and attempt to explain why a proof of renormalizability is still epistemically important in HEP.

and many different theories (in the sense of having different terms in their Lagrangian) will all reduce to a critical surface in the space of Lagrangians on which only renormalizable terms have nonzero coefficients. The same is suspected to hold true for models in quantum field theory, such that a hierarchy of increasingly complicated Lagrangians is expected to obtain at higher and higher energies. These Lagrangians will not contain all and only renormalizable terms, and therefore would require some high-energy cutoff in order to generate empirical predictions. This is why many physicists and philosophers now view quantum field theory as a framework only for effective field theories, and that some new class of theory will be required for a “complete” model of fundamental physics applicable at all energy scales.

This modern view was not developed until after the construction of the electroweak model and QCD, and in fact *could not have been convincingly argued for* without the help of renormalization group techniques. Furthermore, the most convincing argument for a hierarchy of effective field theories—that low energy models of quantum field theory will retain only renormalizable terms—depends crucially on an understanding of the flow from general Lagrangians down to the critical surface of renormalizable interaction terms. The process of proving renormalizability is a two part process: first, one selects a general class of models with similar sets of terms (e.g., Yang-Mills models); second, one must develop appropriate novel techniques for actually proving that this particular class of Lagrangians is actually renormalizable, such as dimensional regularization. This is clearly how Veltman viewed the importance of renormalizability (cf. quote in §4.1), though he would presumably not subscribe to the effective field theory view.

I have shown that the process of arriving at the standard model of particle physics required the coincident development of mathematical tools for dealing with non-Abelian gauge theories and experimental discoveries corroborating the newly discovered properties of the framework. The mathematical developments connected non-Abelian Yang-Mills theories to experiment by demonstrating that renormalizability, asymptotic freedom, and confinement were all properties of QCD, fleshing out the bare Lagrangian form of the candidate model of strong interactions. This case study has provided lessons for the detailed process of theory construction in physics, highlighting the fact that theories are rarely axiomatic systems with a neat set of deductive consequences. Theories like the standard model are instead modular, and rely on key conceptual and mathematical tools that can largely be treated independently. In the case of the HEP, the tools were often constructed by analogy with condensed matter physics. These lessons for theory construction can inform the process of constructing new theories, in particular those attempting to quantize gravity.

## Acknowledgements

I would like to thank Doreen Fraser, Wayne Myrvold, and Marie Gueguen for helpful comments on earlier drafts of this paper. I am also grateful to two anonymous referees, whose suggestions helped to strengthen this paper.



## References

- Bloom ED, Coward DH, DeStaebler H, Drees J, Miller G, Mo LW, Taylor RE, Breidenbach M, Friedman JI, Hartmann GC, Kendall HW (1969) High-energy inelastic  $e - p$  scattering at  $6^\circ$  and  $10^\circ$ . *Phys Rev Lett* 23:930–934, DOI 10.1103/PhysRevLett.23.930, URL <https://link.aps.org/doi/10.1103/PhysRevLett.23.930>
- Bogen J, Woodward J (1988) Saving the phenomena. *The Philosophical Review* 97(3):303–352
- Breidenbach M, Friedman JI, Kendall HW, Bloom ED, Coward DH, DeStaebler H, Drees J, Mo LW, Taylor RE (1969) Observed behavior of highly inelastic electron-proton scattering. *Phys Rev Lett* 23:935–939, DOI 10.1103/PhysRevLett.23.935, URL <https://link.aps.org/doi/10.1103/PhysRevLett.23.935>
- Brown L, Dresden M, Hoddeson L (eds) (1989) Pions to quarks: Particle physics in the 1950s. Cambridge University Press
- Buchholz D, Verch R (1995) Scaling algebras and renormalization group in algebraic quantum field theory. *Reviews in Mathematical Physics* 7(8):1195–1239
- Cao TY (2010) From current algebra to quantum chromodynamics: A case for structural realism. Cambridge University Press
- Cushing JT (1990) Theory construction and selection in modern physics: The S matrix. Cambridge University Press
- Dyson FJ (1949a) The radiation theories of Tomonaga, Schwinger, and Feynman. *Physical Review* 75(3):486–502
- Dyson FJ (1949b) The S matrix in quantum electrodynamics. *Physical Review* 75(11):1736–1755
- Faddeev LD, Popov VN (1967) Feynman diagrams for the yang-mills field. *Physics Letters B* 25(1):29–30
- Feintzeig BH (2017) On theory construction in physics: Continuity from classical to quantum. *Erkenntnis* 82(6):1195–1210
- Feynman RP (1948) Relativistic cut-off for quantum electrodynamics. *Physical Review* 74(10):1430–1438
- Feynman RP (1949) Space-time approach to quantum electrodynamics. *Physical Review* 76(6):769–789
- Fraser D (2018) The development of renormalization group methods for particle physics: Formal analogies between classical statistical mechanics and quantum field theory. *PhilSci-Archive Preprint*
- Fraser D, Koberinski A (2016) The higgs mechanism and superconductivity: A case study of formal analogies. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 55:72–91
- Fraser JD (2017) The real problem with perturbative quantum field theory. *The British Journal for the Philosophy of Science*
- Gell-Mann M, Lévy M (1960) The axial vector current in beta decay. *Il Nuovo Cimento* 16(4):705–726

- Gell-Mann M, Low FE (1954) Quantum electrodynamics at small distances. *Physical Review* 95(5):1300
- Glashow SL (1961) Partial-symmetries of weak interactions. *Nuclear Physics* 22(4):579–588
- Gross DJ, Wilczek F (1973) Ultraviolet behavior of non-abelian gauge theories. *Physical Review Letters* 30(26):1343–1346
- Halvorson H, Müger M (2007) Algebraic quantum field theory. In: *Philosophy of Physics*, Elsevier
- Heisenberg W (1946) Der mathematische rahmen der quantentheorie der wellenfelder. *Zeitschrift für Naturforschung* 1:608–622
- 't Hooft G (1971a) Renormalizable Lagrangians for massive Yang-Mills fields. *Nuclear Physics B* 35:167–188
- 't Hooft G (1971b) Renormalization of massless Yang-Mills fields. *Nuclear Physics B* 33:173–199
- 't Hooft G (1997) Renormalization of gauge theories. In: *The rise of the standard model: A history of particle physics from 1964 to 1979*, Cambridge University Press
- 't Hooft G, Veltman M (1972) Regularization and renormalization of gauge fields. *Nuclear Physics B* 44:189–213
- Kadanoff LP (1966) Scaling laws for Ising models near  $T_c$ . *Physics* 2(6):263–272
- Kinoshita T (2013) Tenth-order QED contribution to the electron  $g - 2$  and high precision test of quantum electrodynamics. In: *Proceedings of the Conference in Honour of the 90th Birthday of Freeman Dyson*, World Scientific
- Landau L, Abrikosov A, Khalatnikov I (1954) “asimptoticheskoe vyrazhenie dlya funktsii grina v kvantovoi elektrodinamike” (“asymptotic expression for the green’s function of a photon in quantum electrodynamics”). In: *Dokl. Akad. Nauk SSSR*, vol 95, pp 1177–1180
- Politzer DH (1973) Reliable perturbative results for strong interactions? *Physical Review Letters* 30(26):1346–1349
- Ruetsche L (2018) Perturbing realism. In: forthcoming in *Realism and the Quantum*
- Schweber SS (1994) *QED and the men who made it: Dyson, Feynman, Schwinger, and Tomonaga*. Princeton University Press
- Schwinger J (1948) Quantum electrodynamics. I. A covariant formulation. *Physical Review* 74(10):1439–1461
- Schwinger J (1949) Quantum electrodynamics. II. Vacuum polarization and self-energy. *Physical Review* 75(4):651–679
- Streater RF, Wightman AS (1964) *PCT, spin and statistics, and all that*. W. A. Benjamin, New York
- Stueckelberg ECG, Petermann A (1953) La normalisation des constantes dans la théorie des quanta. *Helvetica Physica Acta* 26
- Tomonaga SI (1946) On a relativistically invariant formulation of the quantum theory of wave fields. *Progress of Theoretical Physics* 1(2):27–42
- Veltman M (1997) The path to renormalizability. In: *The rise of the standard model: A history of particle physics from 1964 to 1979*, Cambridge University

Press

- Wallace D (2011) Taking particle physics seriously: A critique of the algebraic approach to quantum field theory. *Studies in History and Philosophy of Modern Physics* 42:116–125
- Weinberg S (1967) A model of leptons. *Physical review letters* 19(21):1264
- Weinberg S (1979) Phenomenological lagrangians. *Physica A: Statistical Mechanics and its Applications* 96(1-2):327–340
- Weyl H (1918) Gravitation und elektrizität. *Sitzungsberichte Akademie der Wissenschaften Berlin* pp 465–480
- Wick GC (1938) Range of nuclear forces in yukawa’s theory. *Nature* 142(3605):993
- Williams P (2017) Scientific realism made effective. *The British Journal for the Philosophy of Science*
- Wilson KG (1965) Model hamiltonians for local quantum field theory. *Physical Review* 140(2B):B445
- Wilson KG (1969) Non-lagrangian models of current algebra. *Physical Review* 179(5):1499
- Wilson KG (1971a) Renormalization group and critical phenomena I. Renormalization group and the Kadanoff scaling picture. *Physical Review B* 4(9):3174–3183
- Wilson KG (1971b) Renormalization group and critical phenomena II. Phase-space cell analysis of critical behavior. *Physical Review B* 4(9):3184–3205
- Wilson KG (1971c) Renormalization group and strong interactions. *Physical Review D* 3(8):1818
- Wilson KG (1974) Confinement of quarks. *Physical Review D* 10(8):2445–2459
- Wilson KG (1975) The renormalization group: Critical phenomena and the Kondo problem. *Reviews of Modern Physics* 47(4):773
- Wilson KG (1982) The renormalization group and critical phenomena. *Nobel Lectures, Physics 1981-1990*
- Wilson KG, Kogut J (1974) The renormalization group and the  $\epsilon$  expansion. *Physics Reports* 12(2):75–199
- Yang CN, Mills RL (1954) Conservation of isotopic spin and isotopic gauge invariance. *Physical review* 96(1):191
- Yukawa H (1935) On the interaction of elementary particles. i. *Proceedings of the Physico-Mathematical Society of Japan 3rd Series* 17:48–57