The Twin Origins of Renormalization Group Concepts

James D. Fraser

Please cite published version in *Studies in History and Philosophy of Science*
https://doi.org/10.1016/j.shpsa.2021.08.002

Abstract

This paper traces the origin of renormalization group concepts back to two strands of 1950s high energy physics: the causal perturbation theory programme, which gave rise to the Stueckelberg-Petermann renormalization group, and the debate about the consistency of quantum electrodynamics, which gave rise to the Gell-Mann-Low renormalization group. Recognising the different motivations that shaped these early approaches sheds light on the formal and interpretive diversity we find in contemporary renormalization group methods.

1 Introduction

The renormalization group saw its most famous successes, and established its central place in the toolkit of contemporary theoretical physics, in the 1970s. This paper examines the origins of renormalization group ideas two decades earlier, in the 1950s.

There are a number of motivations for undertaking such a project. The historical development of the renormalization group has been little studied and from the intellectual historian's perspective it makes sense to start at the beginning. Understanding the early renormalization group is a prerequisite for investigating how the work of Kenneth Wilson and others, in the 70s, built on and diverged from, these older approaches.

Furthermore, these questions about the conceptual evolution of the renormalization group are not mere historical curiosities, as they have an important bearing on the foundations of quantum field theory (QFT). A common view amongst physicists, which has also recently been developed in the philosophy of physics literature (Wallace 2011, Fraser 2020), is that Wilson's incarnation of the renormalization group played a decisive role in resolving conceptual problems which had been present since the invention of renormalization techniques. This explanatory claim implies a story of conceptual progress in renormalization theory which can be illuminated by historical investigation.

In addition to preparing the way for this broader historico-philosophical study of renormalization theory, studying the early renormalization group turns out to have more self-contained lessons. When approaching the subject historically, it quickly becomes evident how misleading it can be to speak of "the renormalization group". There are really multiple distinct formal structures that are referred to as renormalization groups in the contemporary literature, as well as multiple interpretations of those structures. This multiplicity is already evident in the 1950s, where we find two independent origin points for renormalization group ideas.

The story, in brief, is this. Stueckelberg and Petermann (1951a, 1953) postulated a group of transformations between different ways of defining the coefficients of QFT perturbation series. Their primary motivation was to address the existence of ambiguities arising in their heterodox causal formulation of perturbative QFT, but their discussion can, in retrospect be seen as elucidating general structural features of perturbative renormalization. Independently, Gell-Mann and Low (1954) invented a trick for obtaining information about the short distance behaviour of QFT propagators based on writing down functional equations involving a "cut-off" scale. In doing so they intervened in a debate that was raging at the time about whether QFTs, and especially quantum electrodynamics (QED), could be consistently formulated outside of perturbation theory. These two strands were connected by Bogoluibov and Shirkov (1955a, 1955b) who popularised the term "renormalization group" in their QFT textbook (Bogoluibov and Shirkov, 1959).

While these two original versions of the renormalization group were entwined together by Bogoluibov and Shirkov we can see each of them as having distinct historical echoes. After elaborating the account just sketched in detail, I conclude by reflecting on how this has fed into contemporary conceptions of the renormalization group. An important moral for historical and philosophical engagement with this tradition is drawn: we must be cognizant of the diverse range of formal structures and interpretations that fall under the rubric of *the* renormalization group.

2 Stueckelberg and Petermann's Renormalization Group

The first appearance of a renormalization group concept in the literature is in the papers of Ernst Stueckelberg and André Petermann (Stueckelberg and Petermann 1951a, 1953), though they, in fact, used the term "groupe de normalisation", or normalization group. Stueckelberg is one of those unfortunate figures in the history of science who is primarily known for anticipating ideas that would later be associated with more celebrated names. These papers with Petermann were more than isolated premonitions, however, and did have an impact on later developments, though admittedly by a circuitous route. Their limited readership at the time (and today) is partially explained by their use of Stueckelberg's unorthodox causal formulation of perturbative QFT (their elliptical prose, idiosyncratic notation, and use of the French language likely also did not help). This approach never penetrated mainstream physics in the English speaking world, but it was further developed by Nicholay Bogoluibov in the Soviet Union. It was Bogoluibov and his student Dmitry Shirkov who took up Stueckelberg and Peterman's "groupe de normalisation" and redubbed it the renormalization group.

The little known causal perturbation theory programme that Stueckelberg initiated thus played a forgotten role in the early development of renormalization group ideas. Our first order of business will be to explain the basic features of this approach. The historical development of causal perturbation theory is chronicled in detail elsewhere (Blum, Fraser and Miller, 2021). Here I will focus on aspects that provide important background for Stueckelberg and Petermann's introduction of their group.

2.1 The Causal Perturbation Theory Programme

A central theoretical problem that had stalked QFT since its inception was the appearance of ultraviolet divergences in perturbation theory. Perturbation theory is a framework for generating approximations, whereby quantities of interest are expanded in a series in an interaction coupling, which is assumed to be small. When this is done naively in the case of QFT, ill-defined integrals appear in the series coefficients. Written in momentum space, these integrals blow up in the region of arbitrarily large momentum—referred to as the ultraviolet regime. Stueckelberg's goal in the 40s was to formulate a perturbative formalism for relativistic quantum theory which was free of this problem. He was thus a competitor of Tomonaga, Schwinger and Feynman, whose work in the same period would go on to form the basis of renormalized perturbation theory as it is practised in mainstream high energy physics today. Henceforth in this section, I will somewhat anachronistically refer to this formalism as conventional perturbative QFT.

A key feature of Stueckelberg's approach which set it apart was its grounding in Heisenberg's S-matrix programme (Heisenberg 1943a, Heisenberg 1943b; see Blum (2017) for a historical treatment). The S-matrix is an operator which maps asymptotic states at $t = -\infty$, interpreted as the 'incoming' particles of a scattering process and represented by states of a free (i.e. interactionless) QFT, to 'outgoing' (and again free) states at $t = \infty$. It is a crucially important object in conventional perturbative QFT. Indeed, following Dyson (1949), it is S-matrix elements for particular scattering processes which QFT perturbation series are primarily used to approximate. Heisenberg's ambitions for the S-matrix had been rather grander than this. In conventional perturbative QFT, the series expansion for the S-matrix is (somewhat informally) derived via the integration of a differential time-evolution equation of the Schrodinger type. Heisenberg believed that quantum theory's reliance on such equations was the root of the ultraviolet divergence problem and hoped that the S-matrix might form the basis of a purely integral formulation of quantum theory which eschewed time-evolution equations entirely.

This meant stepping away from the hitherto unquestioned method of obtaining a quantum theory via the canonical quantization of a classical Hamiltonian theory, which raised the question of how a pure S-matrix theory would actually be constructed. For those interested in developing Heisenberg's vision the key question was how to fix a determinate form for the S-matrix without resorting to the use of a time-evolution equation. Stueckelberg's big idea was that a causality condition could be used to determine the form of a series expansion for the S-matrix.[1] Heisenberg had imposed unitarity and Lorentz invariance on his S-matrix, but Stueckelberg argued that an S-matrix satisfying these conditions could still display behaviour he labelled acausal. Heisenberg had considered an S-matrix of the form:

$$S = e^{i\lambda \int d^4x \phi(x)^4}, \tag{1}$$

purportedly describing a quartic self-interaction of a scalar field $\phi(x)$. This S-matrix has a Taylor series expansion in the coupling constant λ:

$$S = e^{i\lambda \int d^4x \phi(x)^4} = 1 - i\lambda \int d^4x \phi(x)^4 - \frac{\lambda^2}{2} \int d^4x \int d^4y \phi(x)^4 \phi(y)^4 + ... \tag{2}$$

Stueckelberg interpreted these terms as permitting the occurrence of retrocausal processes, in which events in the future influence those in the past.

As Stueckelberg realised, a series expansion of the S-matrix element of a particular set of free incoming and outgoing states could be decomposed into two-point functions—products of the free field operators evaluated at two space-time points. Heisenberg's ansatz for the S-matrix would give rise to two point-functions of the form:

$$D(x-y) = \langle 0| \phi(x)\phi(y) |0\rangle = \int \frac{d^3\vec{p}}{(2\pi)^3} \frac{1}{2\omega(\vec{p})} e^{i\vec{p}.(\vec{x}-\vec{y})} e^{-i\omega(\vec{p})(x_0-y_0)}, \tag{3}$$

appearing in the terms of the series expansion. If $x_0 < y_0$ then the field operator with the later time argument acts on the vacuum first, a situation which Stueckelberg interpreted as a particle being created at y_0 and propagated backwards in time to x_0 (Stueckelberg 1944).

After a few false starts, a proscription for imposing causality on the expansion of the S-matrix was eventually advanced in papers co-authored with Dominique Rivier (Stueckelberg and Rivier 1950a, Stueckelberg and Rivier 1950b). There it was argued that the two-point functions occuring in the series expansion of a causal S-matrix must take a particular form:

$$D_c(x-y) = \theta(x_0-y_0) \langle 0| \phi(x)\phi(y) |0\rangle + \theta(y_0-x_0) \langle 0| \phi(y)\phi(x) |0\rangle \tag{4}$$

$$= \int \frac{d^3\vec{p}}{(2\pi)^3} \frac{1}{2\omega(\vec{p})} e^{i\vec{p}.(\vec{x}-\vec{y})} (\theta(x_0-y_0) e^{-i\omega(\vec{p})(x_0-y_0)} + \theta(y_0-x_0) e^{i\omega(\vec{p})(x_0-y_0)}). \tag{5}$$

Stueckelberg called this the causal function, or causal propagator. Today it is more commonly known as the Feynman propagator, the same type of expressions having been introduced by Feynman in his approach to perturbative QED.

[1]The idea that additional conditions needed to be invoked in order to progress Heisenberg's S-matrix approach was pursued by other physicists of this period such as Max Born and Ralph Kronig—see Blum (2017) for details.

In modern parlance, the field operators in this product are time-ordered, meaning that the operator with the earlier time argument always act on the vacuum first. The acausal processes which Stueckelberg had identified in the expansion of Heisenberg's ansatz for the S-matrix had thus been eliminated.

In Stueckelberg and Rivier's scheme causality is imposed on the series expansion of the S-matrix by adding a correction at each order which ensures that all two-point functions that appear in the series take this form. One starts by specifying the first-order term, S_1, of the general expansion:

$$S = 1 + \sum S_n, \tag{6}$$

which corresponds to selecting an interaction potential for the theory. Requiring S be unitary imposes relationships between the terms of the series. At second order we have, $S_2 + S_2^\dagger = -S_1 S_1^\dagger$, which fixes the hermitian part of S_2. The anti-hermitian part of S_2 is then chosen in such a way that the resulting term contains only causal propagators. Higher-order terms can then be iteratively constructed in the same manner. In this way, a series expansion for the S-matrix is generated without mentioning either a Hamiltonian or a time-evolution equation.

By 1950, of course, conventional perturbative QED was already triumphant, and the Stueckelbergian alternative largely fell on deaf ears. In retrospect, the causal approach was still in a nascent stage of development at this point and substantive issues with Stueckelberg and Rivier's formalism further limited the prospects for wider adoption. Stueckelberg's arguments about the causal behaviour of the S-matrix were poorly communicated and rested on the dubious practice of interpreting individual terms in the perturbation series as corresponding to physical processes, rendering the justification for his condition murky. Furthermore, his formulation of the causality condition was unwieldy, making questions about the behaviour of the resulting series difficult to answer. Stueckelberg and Rivier (1950b) claim to reproduce Schwinger's (1949) second-order expression for the anomalous magnetic moment of the electron, and a paper co-authored the next year with another student, T. A. Green, treated third and fourth-order QED terms in the causal approach (Stueckelberg and Green, 1951). Still, whether Stueckelberg's approach agreed with conventional perturbation theory at all order was, at this point, unclear.

Luckily for us, later iterations of what would become called causal perturbation theory clarified many of these issues, and it will thus be useful to briefly sketch the forward trajectory of the programme Stueckelberg had initiated. The Soviet mathematician turned physicist Nicholay Bogoluibov, who will play an important role later on in our story, was one of the few authors to take up Stueckelberg's ideas in the 50s. He introduced a new formulation of the causality condition which was stated in non-perturbative terms and possessed a more natural physical interpretation (Bogoluibov 1955).[2] Furthermore, Bogoluibov was able to show

[2]Bogoluibov's formulation of the causality condition made use of another concept invented by Stueckelberg: the adiabatic switching function $g(x)$ (for a detailed discussion see Blum, Fraser

that imposing his version of the causality condition on a series expansion for the S-matrix comprised of arbitrary products of field operators reproduces the form of the Dyson series, the expansion of the S-matrix that occurs in conventional perturbative QFT (Bogoluibov and Shirkov 1959). A later mathematical physics tradition, following the work of Epstein and Glaser (1973), developed Bogoluibov's formalism still further, focusing on integrating a rigorous distribution theoretic analysis of the perturbative terms into the causal approach (on which, more in section 2.2). By this time, causal perturbation theory was firmly established as a rational reconstruction of conventional perturbative QFT rather than a rival theory.

There was, however, an important difference between the causal and conventional approach to perturbative QFT when it came to the resolution of the problem of ultraviolet divergences, which feeds directly into Stueckelberg and Petermann's introduction of their renormalization group concept. In conventional perturbative QED, the problem of ultraviolet divergences was tackled via the introduction of a renormalization procedure that works in the following way. The divergent integrals are first replaced by "regularized" convergent integrals; the simplest way to do this being to impose an upper bound, or "cutoff", on the range of integration in momentum space. The fields, masses and charges of the theory, which are now dubbed "bare" quantities, will depend on the value of this cutoff. The theory's dynamics is then rewritten in terms of a new set of renormalized fields, masses and couplings, together with a set of so-called counterterms. In the case of QED, it turned out to be possible to completely remove the part of the perturbative coefficient at each order which diverges as the cutoff goes to infinity by introducing renormalized versions of the field normalization constants, masses and charges appearing in the original Lagrangian (more on the general concept of renormalizability later). When the cutoff is removed, the bare and counterterm parameters go to infinity, but, if this redefinition process has been carried out correctly, we end up with a finite series coefficient multiplying a power of a finite renormalized charge.

and Miller, 2021). $g(x)$ is a smooth function which multiplies the interaction potential of a theory, such that space-time regions where $g(x) = 0$ correspond to a free theory, and those where $g(x) = 1$ correspond to the interaction being fully 'switched on'. Stueckelberg (1951) had introduced the switching function in order to describe scattering processes occurring over finite time periods using an S-matrix (which standardly relates states at asymptotic times), but Bogoluibov realised that it could be used to provide an improved statement of Stueckelbergian causality. Consider two switching functions $g_1(x)$ and $g_2(x)$ which take non-zero values in two finite space-time regions G_1 and G_2, where all of the points in G_1 are in the past with respect to some refernece frame and all of the points in G_2 are in the future with respect to the same reference frame. Bogoluibov's causality condition is then:

$$S(g_1 + g_2) = S(g_2)S(g_1). \tag{7}$$

The interpretation of this statement being that the effect of an earlier period of scattering (in the G_1 region) on the final outgoing state is independent of that of a later period of scattering (in the G_2 region)—that is, retrocausal influence is ruled out. (Note that Bogoluibov (1955) gives an equivalent, but slightly more difficult to read, differential formulation of this causality condition. The integral expression given above appears in later works.)

The causal approach to perturbation theory led Stueckelberg to a conceptually quite different approach to the ultraviolet divergences problem. While ultraviolet divergent terms are so-called because of their behaviour in momentum space, in position space the behaviour of the causal/Feynman propagator at $x - y = 0$ can be seen as the culprit. Stucekberg and Rivier point out that, since causality considerations concern the temporal ordering of events they don't tell us anything about how the propagators behave at these points were the space-time arguments coincide. This led them to state that the products of two-point functions that appear in their perturbation series are in fact ambiguously defined at these problematic coincident points: one can add terms which vanish everywhere except at $x = y$—Dirac delta functions, $\delta(x - y)$, and their derivatives—without violating Stueckelberg and Rivier's (or Bogoluibov's) causality condition. The causal formulation of perturbation theory thus "replaces the problem of eliminating divergences with that of determining the undefined terms of the S matrix" (Stueckelberg and Rivier 1950a, 215, translation).

This might not seem to be a great improvement. As Stueckelberg and Peterman would later write, in light of these ambiguities allowed by the causality condition, "The S-matrix may, therefore, seem at first glance to be completely indeterminate" (Stueckelberg and Petermann 1953, 513, translation). Stueckelberg and Rivier, claimed that the ambiguity could be "partially eliminated" by physical considerations, but this remained another underdeveloped aspect of their work in 1950. Stueckelberg was clearly aware of this deficiency, as the nature and implications of these perturbative ambiguities was the central issue framing his work with Petermann, including the introduction of their "groupe de normalization".

2.2 The "Groupe de Normalization"

Stueckelberg's collaboration with André Petermann, which represents the final phase in the development of his approach to relativistic quantum theory, was apparently triggered by a fundamental realisation about the mathematical machinery needed to properly formalise perturbative QFT—a discovery that would shed new light on the ambiguities in the series expansion that Stueckelberg and Rivier had identified. According to Petermann's recollections:

> "[I was] a recently graduated mathematician working on direct products of generalized functions and on the space on which they can be defined. Stueckelberg, having such a problem with his student T. A. Green, in the work they were doing in S-matrix theory, asked me if I would be interested in working at the Swiss Atomic Energy Commission, in order to deal with this problem in a mathematical way, according to Schwartz, Sobolev and others."[3]

It seems that, circa 1950, Stueckelberg made a connection between the problematic terms appearing in QFT perturbation series and the theory of generalised functions, or distributions, which had recently been developed in pure mathematics

[3] Taken from personal email correspondence between Petermann and Andreas Aste in 2007.

by Laurent Schwartz.

In Schwartz's approach, distributions are defined as maps from a space of test functions to the numbers, a definition which encompasses ordinary functions but also includes so-called singular distributions like the Heaviside and Dirac delta 'functions', which are not functions in the classical sense. The causal/Feynman propagators that appear in causal and conventional perturbation theory are, in fact, also singular distributions, with the singularity occurring at the coincident point $x = y$. From the beginning, Schwartz had emphasised an important difference between distributions and ordinary functions: multiplications and divisions of distributions with overlapping singularities are not generally well-defined. And, it turns out, the problematic terms in QFT perturbation series which give rise to ultraviolet divergences in the conventional approach are products of distributions with overlapping singularities. In contemporary mathematical physics, the problem of ultraviolet divergences and perturbative renormalization is thus often reconstructed as one of giving a proper definition for these products of distributions that appear in the expansion. Stueckelberg and Petermann appear to have been among the first to advance this formulation of the problem.[4]

After publishing two short notes in 1951, which I will come back to later, Stueckelberg and Petermann's project culminated in the paper "La normalisation des constants dans la théorie des quanta" (The normalization of constants in quantum theory) in 1953 which I will focus on here.[5] This is a difficult work. It is long by Stueckelberg's standards but retains the usual density of obscure statements and notational oddities. I will not try to explain all of its intricacies here but focus instead on the line of thinking which directly leads into the introduction of their "groupe de normalization".

It was in this work that the integration of distribution theory into the causal perturbation theory programme came to fruition. They clearly state that ultraviolet divergences arise in the conventional approach because when one starts from a time-evolution equation one ends up writing down products of distributions that are ill-defined. They then go on to develop a procedure for providing a proper definition for the problematic terms which works in the following way. First, the distributions are multiplied by a kinematic factor which compensates for the singularity at coincident points. Products of propagators can then be defined in the same manner as the products of ordinary functions. One then carries out a division of this well-defined product so as to recover the distribution of interest. Consider

[4]The mathematician Georges De Rham, a colleague of Steuckelberg's for two decades at this point at the Universities of both Geneva and Lausanne, appears to have played an important role in transmitting distribution theoretic ideas. De Rham was an early promoter of Schwartz's new theory, inviting him to present his theory in Lausanne in 1948. Stueckelberg likely came across distribution theory, and realised its relevance to QFT, through conversations with De Rham, and if Petermann was already familiar with these ideas from his undergraduate studies, as the above recollections suggest, it must have been from lectures given by De Rham.

[5]In the 1951 proceedings of Swiss Physical Society Stueckelberg and Petermann published a brief paper entitled "The normalization group in quantum theory" which is the first appearance of a renormalization group concept (Stueckelberg and Petermann 1951b).

a product like $D_c(x-y)^2$, which occurs in the series expansion of a scalar field theory. In four space-time dimensions, $D_c(x-y)$ has a singularity like $(x-y)^{-2}$ for $x \approx y$, thus if we write down, $(x-y)^4 D_c(x-y)^2$ we have eliminated the singular behaviour at the origin. The idea is one can then define the series coefficients by dividing this intermediary by $(x-y)^4$.

Stueckelberg and Petermann were a little too ahead of their time here, ploughing on past where the mathematics had been firmly worked out. A systematic theory of the division of distributions was not yet available in 1953 and seems not to have been treated fully in pure mathematics until 1960 in the works of Bernard Malgrange (Malgrange 1960). Stueckelberg and Petermann's claims are stated rather than proven, and indeed it was not until the work of Epstein and Glaser (1973) that the project of treating the products appearing in QFT perturbation series in a fully distribution theoretic way was realised with mathematical rigour. Still, their gist of what they say is confirmed by this later, more systematic work. They state that the division problem for determining the perturbation coefficients does not have a unique solution. Instead, it defines a class of distributions that differ by the addition of Dirac delta distributions, and their derivatives. Consequently, the coefficients of the expansion at each order will take the form:

$$S_n = \int dx_1...dx_n[T_n(x_1,...,x_n) + c_1\delta(x_1,...x_n) + c_2\frac{\partial}{\partial x}\delta(x_1,...,x_n) + ...] \tag{8}$$

where T_n is determined by the causality condition, but the additional terms contain constants $c_1, c_2, ...c_i$ which are entirely arbitrary.

We have thus returned to the earlier conclusion of Stueckelberg and Rivier, that the coefficients of QFT perturbation series contain delta function type ambiguities, but with a few important sophistications. In the distribution theoretic analysis the power of derivatives of the Dirac delta distribution which can be added to the perturbative coefficient at each order is determined by the strength of the singularity of the relevant product—roughly speaking, one cannot make the distribution "more singular" by the addition of these ambiguous terms. We can now see that the ambiguity that arises is not only due to the inapplicability of the causality condition to coincident points, but also flows from the singular nature of QFT two-point functions.[6] The problem of the apparent indeterminacy of the S-matrix does not appear to have been rendered any less severe, however. It is in the course of giving an argument to ameliorate this worry that Stueckelberg and Petermann introduce the "groupe de normalisation".

They first introduce a set of transformations which move us between different ways of fixing the arbitrary constants in the perturbative coefficients of the expansion of the S-matrix:

$$\delta S_n = \sum_i \delta c_i P_i S_n, \quad P_i = \partial/\partial c_i. \tag{9}$$

[6]This explains why these kinds of ambiguities do not arise in non-relativistic quantum mechanics perturbation theory, though it can also be formulated using a causality condition (Scharf 2014). Because the propagators are not singular distributions in this case there is a unique extension of functions given by the causality condition to coincident space-time points.

They claim that these P_i operators satisfy a Lie algebra equation:

$$[P_i, P_k] = -\sum_l L^l_{ik} P_l, \tag{10}$$

and, therefore, act as generators for a transformation group. Again, this is more or less stated rather than proven—later mathematical physicists would be much more careful in establishing the nature of this group structure (see, for instance, Dütch 2020). They then argue that moving between different values of the c_i's via the action of this transformation induces a variation in the dynamical parameters appearing in the series expansion. In QED, for instance, varying the c_i's that appear in the perturbative expansion of the S-matrix is equivalent to varying the mass and charge of the electron which appears in the series. They call this postulated group structure the normalization group rather than the renormalization group, presumably because in the causal approach one avoids initially writing down ultraviolet divergent expressions so one is not renormalizing but simply selecting a normalization when one fixes the values of the c_i's.

Stueckelberg and Petermann take this to resolve the ambiguity problem, but the interpretation of their solution is not entirely clear. One way of reading what they are saying is that, since variations in the c_i constants are equivalent to variations in the dynamical parameters we can eliminate all of the ambiguity by directly measuring, say, the mass and charge of the electron in the case of QED. This would be consistent with the way that the perturbative renormalization procedure was commonly understood at the time. There is, however, another interpretation that is strongly suggested, if not explicitly endorsed, in their article. The fact that they discuss their normalization group alongside the gauge group indicates that they, in fact, want to interpret it as a symmetry transformation of the perturbative S-matrix. Since the effect of any change in the c_i's can be exactly compensated by a change in the masses and charges appearing in the series we can view this compound transformation as leaving the perturbative S-matrix invariant. Many years later, when reviewing his work with Stueckelberg, Petermann would make this interpretation explicit:

> Stueckelberg... noticed that to a change of the arbitrary parameters corresponds a change of the parameters of the theory (couplings, etc.) or, equivalently, that one can find simultaneous changes of the parameters of the theory (couplings, etc.) which compensate the change in the arbitrary parameters, leaving therefore physical quantities invariant. (Petermann 1979)

Some clarifications are needed to understand the full significance of this more interesting reading of Stueckelberg and Petermann's group concept. The reader may have the impression from the discussion thus far that Stueckelberg has merely solved a problem of his own making in a rather convoluted way. After all, the ambiguities in the perturbation series seemed to arise from the heterodox aspects of Stueckelberg's approach; the causality condition and the use of distribution theory. In fact, it is possible to translate Stueckelberg and Petermann's claims into the language of conventional perturbative renormalization and understand them as identifying a general structural property of QFT perturbation series (though they themselves made little attempt to establish such a connection). In the conventional renormalization procedure, counterterms are introduced in order to subtract the divergent part of the perturbative coefficients. During this process, it is in fact also possible to subtract an arbitrary finite part of the coefficient, a fact which does not seem to have been recognised in many of the early papers developing the perturbative renormalization technique. In the contemporary physics literature, the choice of prescription used to assign finite values to the perturbative coefficients is known as a renormalization scheme.

Stueckelberg and Petermann can thus be read as pointing out the invariance of the S-matrix with respect to one's choice of renormalization scheme. Once again, they were ahead of their time here; it would take many decades for the modern concept of a renormalization scheme to crystalise (see section 5 for further discussion). It is interesting to note that the causal perturbation theory programme seems to have played a forgotten role in initiating this process. While the ambiguities Stueckelberg had identified can also manifest in the conventional approach, they are, in some sense, more obvious in the causal perturbation theory approach, whether they are seen as flowing from the causality condition or from distribution theory. We can thus see this as a familiar case of a particular feature of a theory being more transparent in one equivalent formulation than another. Stueckelberg and Petermann's work had essentially no direct impact in mainstream high energy physics, but, as I discuss in section 4, Bogoluibov and Shirkov's adoption of their ideas would represent a step towards the modern concept of a renormalization scheme.

It is worth emphasising what we definitely do not find in the work of Stueckelberg and Petermann, however. They say nothing about a renormalization scale, nor do they connect their renormalization group concept with the scaling behaviour of a QFT in any way. As we shall see shortly, the central application of renormalization group methods in the 50s and 60s was to the question of the asymptotic short distance behaviour of QFT models. This is completely absent from Stueckelberg and Petermann's paper. They took their renormalization group to be important for general foundational, rather than practical, reasons—it shows how perturbative QFT is, in fact, a physically sensible framework, despite the ambiguities which arise in the series coefficients. Adding the apparent lack of calculational application of their group to the esoteric nature of their 1953 paper, it is not hard to see why many of the insights which can be gleaned from it in retrospect were not recognised at the time.

To be fair to Stueckelberg and Petermann, they did put their renormalization group concept to work in one context, using it to develop a notion of renormalizability (though, again, they did not use that term). Renormalizability will be an important concept going forward, so I will conclude this discussion of Stueckelberg and Petermann's work by briefly considering what they say about it. In a note published in Physical Review in 1951 Stueckelberg and Petermann discussed the question of what interaction terms can sensibly be added to the QED lagrangian (Stueckelberg and Peterman 1951a). This was a crucial question at the time. With the empirical success of QED established, the central phenomenological problem in the 50s became the strong nuclear interaction, with various "meson theories", which attempted to describe strongly interacting particles using a field theory in the mould of QED, being considered. Stueckelberg and Petermann treated the question of how the new fields introduced in these theories could consistently interact electromagnetically, concurring with the growing consensus at the time that only scalar, pseudo-scalar and pseudo-vector interaction terms between a meson field and the photon were viable, with vectorial coupling terms being branded physically unacceptable.

Characteristically, they have a heterodox view of why these interactions must be rejected. The dominant view which was emerging at the time was that, in the case of renormalizable interactions, the ultraviolet divergences could be completely removed via the rescaling of a finite number of dynamical parameters, while in the case of these problematic non-renormalizable interactions the ultraviolet divergences problem really was fatal as the divergences could not be systematically removed from the series. In Stueckelberg's approach, of course, ultraviolet divergences never occur in the first place so they more or less had to understand this distinction in a different way. In another 1951 note, published in the proceedings of the Swiss Physical Society, Stueckelberg and Petermann first present a nascent version of their normalization group idea (Stueckelberg and Petermann 1951b).[7] There they say that the normalization group property discussed above holds for non-renormalizable theories, but in these theories, new ambiguities arise at each order in the series corresponding to variations in coupling parameters for new interaction terms.[8] What they identify as pathological about this is that, in the case of abelian gauge theories, this will give rise to an infinite series of higher-order derivative terms. Summing up these terms gives a non-local term in the Lagrangian—a term multiplying field operators at different space-time points. They thus reject non-renormalizable interactions on the grounds that they lead to violations of the physical principles which had been used to construct the perturbation series in the first place.

Stueckelberg and Petermann, yet again, anticipate later conceptual developments in renormalization theory with their emphasis on the fact that non-renormalizable interactions necessitate the introduction of infinitely many dynamical parameters. But the key point from these early discussions of renormalizability which feeds into Gell-Mann and Low's work, to which we now turn, is that one only has to consider rescaling a finite number of dynamical parameters which already appear

[7] In addition to minor notational differences, this brief paper contains no discussion of distribution theory or the origin of the perturbative ambiguities.

[8] In modern developments of the causal perturbation theory approach within mathematical physics the notion of renormalizability is explicated in terms of the strength of the singularities occurring in the products of propagators appearing at each order in the series. In the case of non-renormalizable interactions, the strength of the singularities increases at each order, giving rise to new ambiguities in the coefficients. See, for instance, Scharf (2014).

in the theory if it is renormalizable. As we shall see, Gell-Mann and Low took this fact to provide important information about the theory's short distance behaviour.

3 Gell-Mann and Low's Renormalization Group

In 1954 Gell-Mann and Low published a paper which is also a crucial source of early renormalization group ideas. Like Stueckelberg and Petermann, they did not use that term, in fact, they did not refer to a transformation group at all. The problem they were addressing was quite different from Stueckelberg and Petermann, whose work we can be fairly certain neither was aware of. What we have, then, is something close to a completely independent second origin point for the renormalization group tradition.

Gell-Mann and Low's paper is most naturally situated within the context of a debate that was going on in the early 1950s about the consistency of QED. While they themselves were not particularly passionate participants in this debate, the works of Gunnar Källén (who certainly was) seems to have been their initial jumping-off point, and their paper was soon after its publication used by Bogoluibov and Shirkov to combat the claims about the pathologies of QFT being advanced by Lev Landau and his collaborators. We will thus start by giving some relevant background on this debate.[9]

3.1 The Consistency of QED Debate

The early 1950s was a period of critical reflection on what had really been achieved by the invention of renormalized QED. The renormalization procedure was, at this point, tied to perturbation theory and consequently it remained unclear whether it had solved longstanding worries about the foundations of QFT or merely rehabilitated an approximation technique. As theorists tried to free themselves from the confines of the perturbation series, worries about the high-momentum/short-distance behaviour of QED, and QFT as a whole, quickly recurred.

Perhaps the most ardent pursuer of this line of investigation was Gunnar Källén. In a 1952 paper Källén made the first strides towards translating the renormalization procedure into non-perturbative language using what is now called the Källén-Lehmann spectral representation of the two-point functions (Källén 1952).[10] By inserting a complete set of eigenstates, $|\alpha\rangle$, with momenta p^μ_α, it is possible to relate the two-point propagator of the fully interacting field theory to the causal/Feynman propagator of the free theory. For a scalar field, for instance:

$$\langle\Omega|\, T\phi(x)\phi(y)\,|\Omega\rangle = \int_0^\infty d(M^2)\rho(M^2)D_c(x-y, M^2). \tag{11}$$

[9] A detailed discussion of this debate, which treats many subtleties that I gloss over here, is Blum (forthcoming).

[10] Källén (1952, 1953), in fact, focused on the spectral representation of the interacting field commutation relations. My brief presentation here, focusing on the propagators, is closer to what is found in Lehmann (1954) and Gell-mann and Low (1954), as well as modern textbooks.

Here T indicates that the field operators have been time-ordered (as in equation 4), $|\Omega\rangle$ is the vacuum state of the fully interacting theory, $D_c(x-y, M)$ is the causal/Feynman propagator of a free scalar field of mass M, and $\rho(M^2)$ is the so-called spectral function, given implicitly by,

$$\rho(M^2) = (2\pi)^3 \sum_n \delta(M^2 - p_\alpha^2) |\langle\Omega| \phi(0) |\alpha\rangle|^2, \tag{12}$$

which can be read as describing the density of the theory's eigenstates at a given energy level.

As I described in section 2.1, the perturbative renormalization proceeds by distinguishing a set of bare, counterterm and renormalized parameters. Focusing in on the case of the QED charge, in Dyson's terminology the renormalized electric charge e_r had been related to the bare charge e_0 via a renormalization constant Z_3 which parameterises the counterterms:

$$e_r^2 = Z_3 e_b^2. \tag{13}$$

Z_3 is then chosen so that the counterterms completely subtract the ultraviolet divergent part of the series coefficient at each order. In practice, this meant that Z_3^{-1} was, rather unsatisfactorily, identified with a series consisting of divergent integrals. This particular renormalization constant is, in fact, related to the self-energy of the photon field, and consequently, the spectral representation of the photon propagator can be used to provide a putatively exact definition:

$$Z_3^{-1} = 1 + \int_0^\infty \rho_A(M^2) d(M^2), \tag{14}$$

where $\rho_A(M^2)$ is the spectral function of the photon field $A(x)$, defined in an analogous way to equation 12.

It's worth pausing on the physical interpretation of this construction. The perturbative renormalization procedure was often seen through the lens of vacuum polarization, the possibility for electron-positron creation and annihilation events to screen the true charge of a point particle. In our post-renormalization-group interpretations of QFT, it is common to speak of an 'effective charge', or 'running coupling', which varies with the energy scale of the process under consideration. At this stage, however, physicists were operating with a two-level picture: there was the bare charge, which was understood as the intrinsic charge of a pointlike electron, and the renormalized charge which incorporates the (possibly infinite) effects of vacuum polarization and was identified with the value of the electric charge extracted from long-established experimental measurements of the fine-structure constant. What Källén was doing with his new definition for the renormalization constants was transferring this picture into an exact non-perturbative context. That is, he took the spectral representation of Z_3 to relate the measured charge of the electron to the intrinsic charge associated with an exactly solved QFT model.

Now, many physicists had already accepted that the intrinsic charge of QED was infinite in light of the fact that the bare parameters introduced in the perturbative renormalization procedure go to infinity as the regulator is removed (see, for instance, Tomonaga's Nobel lecture; Tomonaga, 1966). In some corners, however, it was hoped that this was an artefact of perturbation theory and did not reflect the true non-perturbative behaviour. Nowadays it is obvious that a QFT can exhibit ultraviolet divergences in its perturbation series yet have a finite bare coupling constant because asymptotically free theories like quantum chromodynamics are taken to realise this scenario. In the 50s this was much less clear. Exploring whether there was conceptual room to make this distinction was part of what was going on in these debates about the consistency of QED outside of perturbation theory.

For Källén, the spectral representation offered a framework for posing, and potentially answering, the question of what the true behaviour of the bare parameters was. The following year, he took the next step in this programme, giving an argument to the effect at least one of the bare parameters of QED must really be infinite (Källén 1953). The argument proceeded as a reductio ad absurdum. First, he supposed that the bare charge, and the other bare parameters of the theory, are finite. This would imply that the integral in equation 14 converges, which requires that $\rho(k^2) \to 0$ as $k^2 \to \infty$. However, using the assumption that the bare quantities of the theory were finite and taking the first few eigenstates of the spectrum as an approximation, Källén was able to derive a lower bound for the asymptotic behaviour of the spectral function. This meant that the integral could not converge, in contradiction with the initial assumption. Thus, Källén concluded, at least one (though, he suggested, probably all) of the bare parameters of QED must be infinite.

Evaluations of the significance of this result varied. For those who took the divergence of the bare parameters in the perturbative renormalization procedure physically seriously, this result might not have been particularly revelatory. It seemed to scupper hopes that QED might be formulated as a completely finite theory outside of perturbation theory, however. Julian Schwinger evidently considered Källén's argument to be a serious threat to the physical foundations of QED, and QFT more broadly, writing in his 1963 Belfer lecture:

> [O]ver this whole field, over the attempt to make practical calculations in the domain of field theory for the past ten or more years, has lain the dead hand of Källén's dictum that one of the renormalization constants of the theory is infinite. That is, that you could not avoid infinities in any attempt to calculate with the theory. (Schwinger 1963)

For Schwinger, and his followers, it was necessary to find loopholes in Källén's argument if QFT was to be rehabilitated as a viable framework for fundamental physics.[11] For Källén, and his old mentor Wolfgang Pauli, however, this was only the first step in a more ambitious project, which aimed to prove that QED suffered

[11] A series of papers by Schwinger's students, Baker, Johnson and Wiley (1963, 1964, 1967), briefly mentioned in section 4.2, followed this agenda of refuting Källén's proof and demonstrating the finitude of the QED bare parameters. See Blum (forthcoming) for a detailed discussion of this work.

from even more severe physical pathologies than infinite bare parameters (Blum, forthcoming).

Meanwhile, in the Soviet Union, Lev Landau and his collaborators, Alexei Abrikosov, Isaak Khalatnikov, and Isaak Pomeranchuk were also claiming that the non-perturbative behaviour of QED was, in fact, even worse than Källén's 1953 argument suggested. Limited lines of communication with Western physicists had left Soviet physicists on the sidelines during the initial development of perturbative renormalization, but this distance from the centre of the action left them well placed to engage in the critical reception of these new methods taking place in the 50s. Landau, who had been a participant in the earlier struggles with ultraviolet divergences before the war, was apparently initially dismissive of the new renormalized perturbation theory (Ioffe, 2002). Eventually, he relented and joined his former students in developing a different approach to investigating the behaviour of the QED bare parameters (Landau, Abrikosov, Khalatnikov 1954a, 1954b, 1954c, Landau and Pomeranchuck, 1955, Landau, 1955).

Landau's group's starting point was to impose a cutoff Λ on the theory's momentum states, which they interpreted as corresponding to a non-zero radius for the electron and look at how the bare charge of such a cutoff theory, $e_b(\Lambda)$, depends on Λ. Note that, despite making suggestive comments in this direction, they also seem not have a robust concept of an effective charge, and continued to operate with a two-level model of charge renormalization. What they had in mind was a series of theories with different cutoffs and bare charges which give rise to the same "physical" renormalized charge, rather than probing a single theory at different momentum scales and observing different values for the charge.

The method they proposed for determining the bare charge's dependence on the cutoff was based on resumming a portion of the perturbation series. Their approach was inspired by the observation that, if one set up a series expansion in $e_b(\Lambda)$ instead of the renormalized charge, the parts of the series coefficients which diverge as $\Lambda \to \infty$, which appear as logarithms of the cutoff, would become extremely large, spoiling the apparent convergence of the series. Consequently, they propose resumming all of those sub-terms in the expansion which contain logarithms of the cutoff, $\ln(\Lambda^2/k^2)$. In Feynman diagram terms, this amounted to singling out a particular class of diagrams called ladder diagrams, which can, in fact, be systematically resummed. Focusing just on this portion of the series, they were able to write down a closed set of equations for the photon propagator, which as we saw earlier can be related to the renormalization constant Z_3, and thus allows one to get at the QED bare charge.

Using this method Landau's group obtained a troubling formula for the relationship between the bare and renormalized charge:

$$e_b^2 = \frac{e_r^2}{1 - \frac{e_r^2}{3\pi}\ln(\Lambda^2/m^2)}. \tag{15}$$

If we hold e_r fixed and gradually increase the cutoff this expression says that e_b^2 diverges at a finite value $\Lambda = me^{3\pi/2e^2}$, the so-called "Landau pole". After this point e_b^2 becomes negative, and approaches 0 from below, corresponding to the renormalization constant Z_3 actually being negatively infinite. A negative value for the squared coupling was associated with the existence of so-called ghost states—states with negative energy that signal the breakdown of unitarity, which Pauli and Källén (1955) had found in a toy QFT model known as the Lee model (Lee 1954). As a result, you will also find the term "Landau ghost" used to refer to the problematic behaviour identified by Landau's group, though they themselves did not emphasise this issue. What they identified as the most severe consequence of their formula was that holding e_b fixed and taking $\Lambda \to \infty$, the renormalized charge e_r actually goes to zero, no matter what value we assign to e_b. Vacuum polarization appears to be so strong that even an infinite bare charge is rendered completely invisible! Thus, a third and final name for the problem Landau's group had identified is the "Moscow zero".

Clearly, if this result was taken seriously, QED could not be considered a coherent theory of point-interacting fields. Nowadays, physicists tend not to lose sleep over the question of whether QED can consistently describe a possible world down to arbitrarily small length scales because it is viewed as an effective field theory, valid only in a limited energy regime. Physicists in the 50s were, of course, well aware that QED was not a fundamental theory. The real significance of these arguments was that QED stood in as a proxy for QFT as a whole. Landau's group claimed that QED only made sense if a cutoff was imposed below the Landau pole, and further argued that in the case of meson QFTs, which were being developed at the time, the pathologies appeared at a much lower momentum scale, indicating that they have no domain of applicability at all. They thus concluded that QFT had reached the end of its usefulness with renormalized perturbative QED. Källén's attitude towards field theory was more nuanced (and cryptic) but the aim of his investigations was also to probe the limits of QFT in order to inform future theorising.

Chronologically, Gell-Mann and Low's paper "On Quantum Electrodynamics at Small Distances", sits between Källén (1953) and the papers of Landau's group. While it clearly fits into this debate about the true values of the QED bare parameters, Gell-Mann and Low appear not to have had a big picture view of the foundational status and future prospects of QFT, and never would return to the topic. Where Källén and Landau were motivated by foundational questions and developed new techniques in the course of trying to answer them, for Gell-Mann and Low it may have been the other way round: they may have noticed a feature of QED perturbation series which seemed to shed light on the asymptotic behaviour of the fully interacting propagators, without taking much of a stance on the significance of this problem.

3.2 Gell-Mann and Low "On Quantum Electrodynamics at Small Distances"

Gell-Mann and Low's collaborative relationship began when they shared an office together as postdocs at the Institute for Advanced Study in Princeton in 1950. There, the pair co-authored a paper which provided a derivation of the Bethe-Salpeter equation, a previously postulated equation describing a QFTs two-particle bound states (Gell-Mann and Low, 1951). What is interesting about this paper for our purposes is that their derivation proceeds by inserting a complete set of eigenstates into the interacting four-point function, in a very similar way to Källén's derivation of the spectral representation. This lends credibility to Gell-Mann's claim that he independently discovered the spectral representation of the propagators around this time (Gell-Mann, 1987).

Plausibly then, Gell-Mann and Low obtained the spectral representation of the interacting propagators, and like Källén, recognised that it could be used to frame questions about the behaviour of the QED bare parameters outside of perturbation theory. The frequent references to Källén's work, and the fact that they relate their conclusions to Källén's 1953 argument for infinite bare parameters, suggests that his papers may have been the trigger for their second project together, or at least strongly influenced its final form. We know that Gell-Mann and Low sent Källén an early draft of the paper, though unfortunately the correspondence is lost—Källén reported this to Pauli, and Gell-mann and Low (1954) acknowledge discussion with Källén in a footnote.[12]

If they were following on from Källén's work, their approach was, conceptually and technically quite different. Unlike Källén, they focus specifically on the bare charge of QED and frame their discussion around the concept of vacuum polarization. They start, in their introduction, by pointing out that perturbative QED implies the existence of corrections to the classical Coulomb potential associated with charged bodies, and that the deviations are strongest at very short distances. They then say something interesting about the physical interpretation of this vacuum polarization correction:

> [A]s we inspect more closely and penetrate through the [electron-positron] cloud to the core of the test body, the charge that we see inside approaches the bare charge q_0, concentrated in a point at the center. (Gell-Mann and Low, 1301)

Here they seem to posit a series of intermediate values between the renormalized and bare charge. Indeed Gell-Mann and Low are often credited with first introducing the notion of an effective scale-dependent charge (see, for instance, Wienberg 1981). As I will shortly point out, however, their notation still bears

[12]Interestingly, Källén tells Pauli that Gell-Mann and Low were trying to prove the inconsistency of QED, which could indicate that the stated conclusion of the earlier draft Källén read was quite different from the published version, which certainly does not claim to have shown that QED is inconsistent, or, indeed, has infinite bare parameters. This could also just be a rhetorical flourish (or mistake) on Källén's part, however.

the mark of the two-level model of renormalization, and their discussion cannot be neatly mapped onto contemporary presentations of the "running coupling" within perturbative QFT as they do not yet have a concept of a renormalization scale.

In any case, unlike Källén, they clearly see the problem of getting at the bare charge as being one of determining the short distance scaling behaviour of the theory. They start by giving the spectral representation of the interacting photon propagator (which I will denote D_r as they associate it with the sum to all orders of the perturbative expansion written in terms of the renormalized charge e_r) and relating its Fourier transform to the modified Coulomb potential:

$$V(r) = \frac{qq'}{(2\pi)^3} \int d^3k e^{i\vec{k}.\vec{r}} D_r(k^2, e_r^2). \tag{16}$$

For Gell-Mann and Low, then, the true value of the QED bare charge ultimately depends on how $D_r(k^2, e_r^2)$ behaves as the external momentum, k, goes to infinity. (Gell-Mann and Low also treat the asymptotic behaviour of the electron propagator, but since the techniques they develop can be fully illustrated by their treatment of the photon propagator, which they clearly see as more physically significant, I will restrict my attention to that case.)

The approach to this problem which Gell-Mann and Low develop also departs from Källén by being based on the perturbative expansion.[13]. In order to get at the full photon propagator of the theory, they want to relate it to what they call the "cutoff propagator" that appears in perturbation theory. There is some difficulty interpreting what they have in mind here. The first step of the renormalization procedure, recall, is to regularize the ultraviolet divergent integrals. The original way of doing this was to somehow "cutoff" the large momentum part of the integral. Gell-Mann and Low initially discuss a regularization procedure originally introduced by Feynman, which can straightforwardly be understood as suppressing the high momentum contributions in ultraviolet divergent integrals appearing in the series. However, they go on to develop a more sophisticated procedure based on John Ward's systematic treatment of the renormalization of QED (Ward, 1951). Ward had made explicit what had often gone unsaid in previous presentations: that the subtraction of ultraviolet divergences is carried out in such a way that, at zero external momentum, the renormalized propagator is set equal to the propagator of the free theory. Gell-Mann and Low point out that we can also set the renormalized propagator equal to the free propagator at some other momentum scale, λ, which they continue to call a "cutoff".

Formally, the momentum scale they have introduced is essentially what we now call a renormalization scale (more on this notion shortly when we come to discuss the work of Bogoluibov and Shirkov). However, the fact that they continue to call λ a cutoff indicates that they are not interpreting it in the way we do today. Landau's group, remember, had also introduced a cutoff, interpreting it physically as a finite

[13]Källén had been trying to get away from the perturbation series entirely, which he viewed as an unreliable starting point, leading him to criticise Gell-Mann and Low's paper in a review for Zentralblatt der Mathematik (Källén 1954).

radius of the electron. They had thus imagined different values of the cutoff as corresponding to different non-local theories. There is no indication that this is what Gell-Mann and Low have in mind, however. More often, in perturbation theory, the cutoff functions were understood as calculational intermediates with no physical significance. This seems to be the way they view their cutoff photon propagator D_λ, which they identify with the perturbative expansion of the photon propagator to all orders calculated with their λ "cutoff". They continue to view D_r as the primary object of interest and seem to view D_λ as playing a formal role in their technique for probing its large momentum behaviour. The two-level view of renormalization thus continues to make itself manifest in Gell-Mann and Low's notation, even as they lay the foundations for a more sophisticated picture.

The starting point for their method is the following expression they write down connecting the "cutoff" and renormalized photon propagator:

$$D_\lambda(\lambda^2, k^2, m^2, e_\lambda^2) = Z_{3\lambda}(\lambda^2, m^2, e_r^2) D_r(k^2, m^2, e_r^2). \tag{17}$$

Where m is the mass of the electron, k is the external momentum, λ is this new "cutoff" scale, and e_λ is the charge appearing in the "cutoff" perturbation series. This relation, they say, holds to all orders in perturbation theory because QED is renormalizable. Their idea here seems to be that, because the interaction is renormalizable we know that the "cutoff" dependence in D_λ can be completely removed via the rescaling of the charge alone (via the constant $Z_{3\lambda}$). They then argue that, at very small length scales, or equivalently large momentum, we can set, $|k^2|, \lambda^2 \gg m^2$, neglecting the mass dependence of the cutoff propagator, and write, $D_\lambda = d_\lambda(k^2/\lambda^2, e_\lambda^2)/k^2$. The renormalized propagator can similarly be written, $D_r = d_r(k^2/m^2, e_r^2)/k^2$. Finally, they argue that the renormalization constant is given by, $Z_{3\lambda}^{-1} = [d_r(k^2/m^2, e_r^2)]_{k^2=\lambda^2}$. Employing these kind of dimensional arguments they reach a pair of functional equations:

$$d_\lambda(k^2/\lambda^2, e_\lambda^2) = \frac{d_r(k^2/m^2, e_r^2)}{d_r(\lambda^2/m^2, e_r^2)}, \tag{18}$$

$$e_\lambda^2 = e_r^2 d_r(\lambda^2/m^2, e_r^2), \tag{19}$$

Gell-Mann and Low understand these equations as expressing a constraint on the large momentum dependence of the perturbative expansion to all orders which flows from the renormalizability of the interaction and dimensional analysis.[14]

[14]Note that there is an issue here about what it really means to equate D_r and D_λ with the sum of an "all orders" perturbation series. There was a growing consensus in the early 50s that, even after renormalization, QED's perturbation series likely did not converge. (Incidentally, André Petermann contributed to this literature while he was working with Stueckelberg, Petermann (1953)). Gell-Mann and Low address this issue only briefly, noting: "[T]he convergence properties of this series are unknown. We have assumed throughout, however, that it defines a function which satisfies the same functional equations that we have derived for the series" (Gell-mann and Low, 1954, 1308). They thus assume that it will be possible to assign some kind of the sum to the series even if it turns out not to converge. Even granting this, though, a question remains about whether a sum assigned to a quantities perturbative expansion can be equated with its value in an exact solution of the model. In my view, questions of this kind remain pertinent in contemporary implementations of the perturbative renormalization group.

These functional equations admit different classes of solutions, which Gell-Mann and Low relate to various possible short distance behaviours for QED. They write the general solution to their functional equations in a few forms, but the most useful for relating their discussion to the later evolution of the renormalization group is the following integral equation:

$$k^2 \to \infty : \qquad \ln \frac{k^2}{m^2} = \int_{q(e_r^2)}^{e_\lambda^2(\lambda^2 = k^2)} \frac{dx}{\psi(x)}. \tag{20}$$

They read this asymptotic condition in the following way. As $k^2 \to \infty$ the left-hand side goes to infinity. $e_\lambda(\lambda^2 = k^2 = \infty)$ is identified with the bare charge of the theory, so the question, as they see it, is at what value of its upper limit the integral on the right-hand side diverges. This is determined by the behaviour of the unknown function $\psi(x)$.[15]

They distinguish two possible scenarios for QED's short distance behaviour. The integral may diverge when e_λ takes some finite value, corresponding to the existence of a finite bare charge for QED. This occurs if $\psi(x)$ has a zero at some finite value $x_0 = e_b^2$. Gell-Mann and Low stress that, in this case, the finite value of the bare charge is independent of the renormalized charge e_r. What they are anticipating here is the possibility of what is nowadays called an ultraviolet fixed point. In an appendix, Gell-Mann and Low define their ψ-function in the following way:

$$\psi(e_\lambda^2) = \frac{\partial(e_\lambda^2)}{\partial(\ln{(k^2/\lambda^2)})}. \tag{21}$$

A reader familiar with contemporary field theory will thus recognise it as a notational variant of what is now called the β-function.[16] The β-function describes how the expansion parameter depends on the renormalization scale, and thus characterises the scaling behaviour of the theory. A zero in the β function is taken to correspond to a fixed point in the "flow" induced by the renormalization group transformation, meaning that the theory is invariant under further scale transformations. In the ultraviolet regime, this implies that the theory's behaviour is asymptotically finite and scale-invariant. Gell-Mann and Low, of course, do not have this picture. In fact, they do not attach a clear physical interpretation to their ψ-function. Formally, however, there is a direct continuity between their discussion and later characterisations of finite asymptotic scaling within perturbative QFT.

The other possibility they consider is that the integral in their asymptotic condition diverges only when the upper limit is infinite, corresponding to an infinite bare charge. They calculate a perturbative approximation for $\psi(x)$ to second-order and

[15]The lower limit of integration, $q(e_r^2)$ is also an unknown function, however, Gell-Mann and Low comment that it is very well approximated by e_r^2 and does not play an important role in determining the asymptotic behaviour.

[16]The β-function is usually defined: $\beta(e(\mu)) = \partial e(\mu)/\partial(\ln(\mu))$, where μ is the renormalization scale. This notation originates in Symanzig (1970).

state that this is the scenario indicated by perturbation theory. However, they imply that they do not consider this to be decisive. They note that Källén's (1953) argument leaves open the possibility of a finite bare charge if one of the other renormalization constants is infinite. They thus implicitly suggest that effects that are invisible within perturbation theory may still give rise to a finite bare charge, and in fact extend this conclusion to other renormalizable field theories, such as the meson theories which were being investigated in the 50s. They say nothing explicit, however, about what the open possibility of a finite or infinite interaction coupling means for the scope of QFT, leaving the foundational upshot of their discussion unspecified.

This attempt to map out the possible short distance behaviour of a QFT model would ultimately turn out to be extremely influential. With hindsight, we can see that Gell-Mann and Low's taxonomy contains some important gaps, however. What they appear to have missed is that inserting their first order perturbative approximation, $\psi(x) \approx (1/12\pi)x^2$, into equation (19) does not give rise to an integral which diverges as $e_\lambda \to \infty$ but one that converges. This corresponds to e_λ going to infinity for some finite value of k^2, i.e. to the existence of a Landau pole. The work of Landau's group had not appeared when Gell-Mann and Low were writing their paper, but they could very well have reached a similar conclusion about the ultraviolet behaviour suggested by QED perturbation theory within their own framework. Bogoluibov and Shirkov would later add the Landau pole scenario to Gell-Mann and Low's list of potential short distance behaviours for QED.

A second, entirely excusable, omission of their discussion is the possibility of asymptotically free theories—theories whose interaction coupling vanishes at short distances. Returning to Källén's spectral representation of the renormalization constants (equation 14), one can see that, if the spectral function is positive definite, Z_3 can only take a value between 0 and 1. This seems to imply that vacuum polarization can only screen the intrinsic value of the charge, and the renormalized charge cannot be larger than the bare charge. Källén had taken the positive definiteness of the spectral function to be a necessary feature of a physically sensible QFT and thus believed that this statement about the relative values of the bare and renormalized charge was generic. Following this view, Gell-Mann and Low assumed that the ψ-function is positive definite. In a theory where $\psi(x)$ takes negative values the limits of integration on the right-hand side of their asymptotic condition can be exchanged, opening up the possibility of a bare charge which is smaller than the renormalized charge. It was not until perturbative calculations of the β-function of non-abelian gauge theories in the 70s indicated a zero bare charge that this scenario was even countenanced, however (t'Hooft, 1999). It was only then that the true breadth and power of the classificatory scheme for which Gell-Mann and Low had laid the groundwork was fully appreciated.

4 Bogoluibov and Shirkov's Renormalization Group

We now have two papers on the table, Stueckelberg and Petermann (1953) and Gell-mann and Low (1954), which appear to develop very different ideas in service of very different aims. I suspect, in fact, that most QFT experts of this period would have seen little connection between them. There was one figure, however, who was uniquely placed to synthesise these two lines of thought. Nicholay Bogoluibov was one of the very few physicists to take up Stueckelberg's causal perturbation theory approach in the 50s. As was discussed in section 2, he developed an improved formulation of the Stueckelberg's perturbative causality condition (Bogoluibov 1955), and presented the causal approach in a much more comprehensible way in a series of papers (Bogoluibov and Shirkov 1955c, 1955d), and finally a textbook (Bogoluibov and Shirkov 1959)[17], co-authored with his student Dmitri Shirkov. He was thus perhaps one of the only physicists in this period who carefully read, and made some sense of, Stueckelberg and Petermann (1953). Being in Moscow in the 1950s, he also had a front-row seat for the hubbub surrounding Landau's claims about the pathologies of QFT.

It was at the confluence of these two intellectual streams that the term "renormalization group" was finally coined. I will start by discussing how this synthesis took place, and what Bogoluibov and Shirkov contributed in the process. I will then consider the puzzle of why this newly minted renormalization group concept did not, in fact, have a major impact in the decade following Bogoluibov and Shirkov's work.

4.1 Bogoluibov and Shirkov's Synthesis

In 1955 Landau presented his group's claims about the inconsistency of QED at a conference on "Quantum Electrodynamics and Elementary Particle Theory" held at the Lebedev Institute in Moscow.[18] This caused quite a stir at the time, and owing to Landau's considerable influence, his anti-QFT interpretation of their results would shape Soviet high energy physics for many years to come.

Bogoluibov was naturally placed to act as a counterpoint to this QFT-skeptic movement in Soviet physics. Where Landau spent the 50s shooting down QFT, Bogoluibov had been trying to shore up its foundations. As well as developing Stueckelberg's causal approach, which he took to be a mathematically and

[17] Originally published in Russian in 1957 and translated into English in 1959.

[18] With the death of Stalin in 1953 it became easier for scientists to travel between the Soviet Union and the West, and a number of foreign field theorists were invited to this conference. As it happens, Källén was the only Western physicist who attended, and debated the fate of QED with Landau there—he was sharply critical of their approach to the subject, due to its basis in the perturbative expansion. In 1956 another high energy physics conference took place in Moscow which many more foreign field theorists attended, including Gell-Mann. Gell-Mann and Landau apparently also discussed his group's claims about the behaviour of the bare charge, with Gell-Mann also being unconvinced (presumably for similar reasons to Bogoluibov and Shirkov).

conceptually superior way of setting up perturbative QFT, he wrote on other fundamental topics, such as the Schwinger-Tomonaga equation (Bogoluibov 1951a, 1951b), and, with Parasyuk developed a rigorous approach to demonstrating the renormalizability of a QFTs perturbative expansion (Bogoluibov and Parasyuk, 1957). As a result, Bogoluibov and his students came to form a pro-QFT school within Soviet physics which rivalled the dominant Landau school.

According to Shirkov's recollections, in the immediate aftermath of the 1955 conference, he and Bogoluibov set about scrutinising Landau's results and investigating the consistency of QED issue themselves. And it was in this context that the Stueckelberg-Petermann and Gell-Mann-Low concepts were fused:

> Shortly after the conference at the Lebedev Institute, Alesha [Alexei Abrikosov] told me about Gell–Mann and Low's article, which had just appeared. The same physical problem was considered in this paper, but, as he put it, it was complex to understand and they had not succeeded in combining it with the results obtained by their group. I looked through the article and presented my teacher with a brief report on the methods and results... The scene that followed my report was quite impressive. On the spot, N.N. [Bogoluibov] announced that Gell–Mann and Low's approach was correct and very important: it was a realization of the normalization group (la groupe de normalisation) discovered a couple of years earlier by Stueckelberg and Petermann in the course of discussing the structure of the finite arbitrariness in the matrix elements arising upon removal of the divergences. (Shirkov, 1996, 4)

In what sense was Gell-Mann and Low's approach a "realization" of Stueckelberg and Petermann's normalization group concept? Making this connection required some reinterpretation of the line of argument which appears in Gell-mann and Low (1954).

Bogoluibov and Shirkov (1955a, 1955b) interpreted the momentum scale, λ, which Gell-Mann and Low had introduced, not as a cutoff, but as an arbitrary scale at which the perturbative expansion of the interacting photon propagator is normalized. In other words, they understood the momentum scale as entering not during regularization but during the renormalization phase, in which a prescription is defined to assign finite values to the perturbative coefficients. They argue that different choices of λ corresponded to physically equivalent definitions of the coefficients and expansion parameters appearing in the series and thus equivalent definitions of the fully interacting propagators.[19] The freedom to fix this scale is embodied in a group of transformations that move us between different values of λ:

$$D_\lambda \to D_{\lambda'} = z_3(\lambda, \lambda')D_\lambda, \quad e_\lambda \to e_{\lambda'} = z_3^{-1}(\lambda, \lambda')e_\lambda. \tag{22}$$

It is this group of transformations that is finally dubbed the renormalization

[19] Note that Bogoluibov and Shirkov also do not address how the non-convergence of the perturbation series affects their argument (see footnote 14).

group. Stueckelberg and Petermann (1953), charitably interpreted, had argued that moving between different values for the arbitrary constants which appear in the expansion coefficients could be compensated by a redefinition of the dynamical parameters appearing in the series. Bogoluibov and Shirkov state that the freedom to normalize the expansion of the propagator at different scales λ is a manifestation of this property.

They then reconstruct Gell-Mann and Low's analysis of the short distance behaviour of QED starting from this transformation property. They obtain an analogous set of functional equations in the limit of arbitrarily large momentum, with an important notational difference. Gell-Mann and Low had maintained a distinction between their renormalized propagator D_r and the cutoff propagator D_λ, apparently attaching physical significance only to the former. This distinction now evaporates. Gell-Mann and Low's D_r is just the propagator obtained with $\lambda = 0$, but all values of the momentum scale are on a par physically on Bogoluibov and Shirkov's reinterpretation. While Bogoluibov and Shirkov do not use the term, they clearly come much closer to the modern concept of a renormalization scale. In contemporary convention perturbative QFT, the renormalization scale is typically characterized as an arbitrary momentum scale at which the subtraction of ultraviolet divergences is carried out. In the causal perturbation theory approach, as we saw, one never writes down ultraviolet divergent expressions or introduces a regulator, so Bogoluibov and Shirkov were more or less forced to reinterpret Gell-Mann and Low's talk of a "cutoff" scale. Again, we see that the causal perturbation theory programme played a forgotten role in drawing attention to perturbative ambiguities which, while also manifest in the conventional approach, had often gone unnoticed. In the following decade, Bogouibov and Shirkov's insight that one could, in principle, impose renormalization conditions which set the photon propagator equal to the free propagator a non-zero external momentum seems to have been quietly taken on board by practitioners of conventional perturbative QFT, paving the way for the modern concept of a renormalization scale.

Bogoluibov and Shirkov also suggested that their renormalization group potentially had a broader significance than just giving information about short-distance structure. They emphasised that functional equations of the kind Gell-Mann and Low wrote down hold not only in the limit of arbitrarily large external momentum but in all regimes since their existence expresses a general structural feature of QFT perturbation series. They also consider the long-distance, or infrared, asymptotic limit, pointing out that analogous dimensional arguments can be used to obtain solvable functional equations in this regime as well (Bogoluibov and Shirkov, 1956). This part of their work would not be picked up by their immediate successors (though nowadays renormalization group methods are, indeed, also used to probe infrared structure). But the fact that Bogoluibov and Shirkov are looking to expand the range of useful applications of their renormalization group concept is significant. As I discuss in the next section, adapting the renormalization group concept to new problems will be an essential precondition for the explosive growth of renormalization group methods in the 70s.

When it comes to the short distance asymptotics problem, an important contribution of Bogoluibov and Shirkov's work is that they relate Gell-Mann and Low's framework with the claims of Landau's group. Gell-Mann and Low, recall, had missed the possibility of a Landau pole, instead only distinguishing between the possibility of a finite and infinite bare charge. Bogoluibov and Shirkov explained how Landau's analysis of QED could, in fact, be fitted into Gell-Mann and Low's framework. They argued that we can understand Landau's resummation approach as computing a perturbative approximation of Gell-Mann and Low's ψ-function. Resumming the "leading logarithms" appearing in the series was equivalent to taking the first-order approximation to $\psi(x)$, and calculating higher-order corrections amounts to resumming higher powers of $\ln(k^2/\lambda^2)^n$. Bogoluibov and Shirkov thus added the Landau pole behaviour to Gell-Mann and Low's categorization of possible short distance behaviours and pointed out that it actually this third option which was suggested by a perturbative approximation of the ψ-function.

This reconstruction of Landau's approach allowed Bogoluibov and Shirkov to identify its shortcomings, however. They point out that a first-order approximation to the ψ-function will only be valid if e_λ is small. But the whole point of Landau's argument is that e_λ will become very large and go to infinity. Landau's group, on Bogoluibov and Shirkov's reading, were therefore guilty of extrapolating a small value approximation outside of its region of applicability. In addition to this point about the limitations of a purely perturbative treatment of asymptotic behaviour, Bogoluibov and Shirkov give an informal argument to the effect that, supposing the bare charge is infinite, introducing a non-local cutoff in the way that Landau's group had suggested may actually give rise to the misleading appearance of a Landau pole. They thus conclude that, when it comes to the asymptotic behaviour of QED, "it is dangerous to make on the basis of approximations any sort of conclusion with respect to the state of affairs in the exact problem" (Bogoluibov and Shirkov, 1959, 529).

The official Bogoluibov and Shirkov line thus seems to have been one of agnosticism about the fate of QED. Given the theoretical tools available at the time it was simply impossible to tell whether QED, or any other QFT for that matter, had finite asymptotic behaviour or something more physically problematic (they do not discuss Källén's non-perturbative arguments). Pragmatically, therefore, it made sense to further develop and refine QFT rather than abandon it. This attitude fits well with the research programme that Bogoluibov and his collaborators pursued during the 50s, which as I said was characterised by modest attempts to shore up the foundational standing of the theory. Bogoluibov and Shirkov's textbook, "An Introduction to the Theory of Quantum Fields", published in Russian in 1957 and translated into English in 1959 collated much of this work. The penultimate chapter on the renormalization group seems to have been the primary line of transmission for the ideas from the 50s we have been surveying into the following decade. All of the articles I have been able to find which continue this tradition in the 60s both use the term "renormalization group" and immediately cite Bogoluibov and Shirkov's textbook. Important figures for later developments, like Wilson and Weinberg, both recall learning of the renormalization group concept from this source as young theorists.

While Bogoluibov and Shirkov's textbook established a place for the renormalization group in the QFT lexicon, it did not establish its centrality. I will now consider the puzzlingly slow uptake of the newly minted renormalization group concept in the following decade.

4.2 Why the Quiet Decade?

The renormalization group remained a fairly marginal concept throughout the 1960s. A bibliographical search for abstracts containing the term "renormalization group" gives a low figure of the order of 10 per year during the 60s which increases by an order of magnitude in the 70s. There is an obvious historical puzzle here: why did it take so long for the renormalization group to rise to its contemporary prominence? This is a wide-ranging question, which stands in need of more careful study, but I will offer some hypotheses.

A major factor behind the weak reception of the 50s renormalization group work was surely its close ties to a class of models which most theorists were desperately trying to get away from by the 60s: renormalizable perturbative QFTs. In the early 50s, there was some hope that it might be possible to describe strongly interacting particles via a formalism that was closely analogous to perturbative QED, but this looked increasingly unlikely as time went on. Meson field theories failed to reproduce the startling empirical successes of QED, and perturbation theory, with its assumption of weak coupling, seemed to be an inherently inappropriate framework for approaching strong nuclear processes. While Gell-Mann and Low (1954) had been trying to probe QED's exact short distance behaviour the method they had developed was intrinsically tied to the perturbative expansion. In a sense, this became more evident with Bogoluibov and Shirkov's synthesis, as the Stueckelberg-Petermann renormalization group concept was all about ambiguities that arise in the course of setting up QFT perturbation series. Furthermore, Gell-Mann and Low had styled their method as being fundamentally based on the concept of perturbative renormalizability, and Bogoluibov and Shirkov had quietly continued to restrict their attention to renormalizable theories. It was thus difficult to see how renormalization group ideas could be salvaged from the now faltering perturbative formalism.

In fact, many of the approaches to strong interactions which rose to prominence in the 60s, such as the S-matrix programme championed by Jeffrey Chew, moved away from QFT entirely. The renormalization group, which had been associated with a cautiously optimistic attitude towards the consistency of QFT question, was likely damned by association in the eyes of many. When Bjorken and Drell dedicated the final subsection of their 1965 QFT textbook to the renormalization group they no doubt further solidified its place in the physics canon, but their final words on the subject are telling:

> "[C]onclusions based on the renormalization group arguments concerning the behavior of the theory summed to all orders are dangerous and must be viewed with due caution. So is it with all conclusions from local relativistic field theories." (Bjorken and Drell, 1965, 376).

This ambivalent attitude towards early renormalization group methods, conditioned by an ambivalent attitude towards QFT, seems to typify the dominant community opinion during this period.

Many of those theorists who did engage with renormalization group ideas in the 60s were QFT loyalists, of one kind or another, the most important being Kenneth Wilson. Wilson's thesis, completed in 1961 under the supervision of Gell-Mann,[20] was an early attempt to employ renormalization group ideas outside of perturbation theory, carrying out a Gell-Mann-Low style analysis of the large momentum limit of the Low equation, a non-perturbative result in fixed source nuclear field theories (Wilson, 1961). Ultimately rejecting Chew's S-matrix approach, Wilson continued to hold the minority view that strong nuclear physics must be underwritten by a QFT. Yet even someone like Wilson, who was committed to QFT and considered the renormalization group work of the 50s to be extremely important, struggled to advance these ideas during the 60s. His papers in the early 70s would finally liberate renormalization ideas from the confines of perturbation theory, and indeed from QFT itself, an achievement that played a decisive role in the explosion of interest in the renormalization group in the 70s (Wilson 1971a, 1971b). This was no straightforward generalization of ideas from the 50s, however. As we will discuss further in the next section, Wilson's renormalization group was something quite different, both formally and conceptually, from what had come before.

A more typical continuation of the renormalization group tradition was the work of Karl-Erik Eriksson. Joining the Cern theoretical division in 1959, Eriksson collaborated with our old friend André Petermann, now a member of the theory division there (Eriksson and Petermann, 1960), and went on to write a series of papers clarifying aspects of Bogoluibov and Shirkov's renormalization group, focusing especially on the connection with the resummation of kinematic logarithms (Eriksson 1961, 1963a, 1963b). The phenomenological background of this work was the precision testing of QED being carried out at Cern, in which context perturbative QFT continued to reign supreme. The fact that publications employing the renormalization group were clustered in this area further supports the point that the intrinsically perturbative nature of the 50s renormalization group limited opportunities for wider adoption. This situation changed radically when the asymptotic freedom of quantum chromodynamics (QCD) was established, meaning that weak coupling approximations were valid at high energies. Parallel to Wilson's efforts in developing a non-perturbative version of the renormalization group, perturbation theory had, surprisingly, returned to centre stage in the guise of perturbative QCD, surely another crucial factor in the resurgence of renormal-

[20] Though apparently, Gell-Mann had relatively little input on Wilson's project. Wilson and Gell-Mann both, somewhat surprisingly, report that personal interactions between them played no role in the transmission of renormalization group ideas.

ization group ideas in the 70s.

Another reason for the lacklustre response to renormalization group methods in the 60s was their weak track record when it came to solving concrete problems. The Gell-Mann-Low and Bogoluibov-Shirkov work had shown, to those that understood it, that Landau's conclusions about the generic pathological short distance behaviour of QFTs were too quick. They had not succeeded in giving much positive information about short-distance structure, however. In the 60s some theorists continued to pursue this problem, in many cases departing from the renormalization group approach. As a PhD student of Wilson's, Roman Jackiw was given the thesis problem of applying renormalization group methods to the short distance behaviour of the QED vertex function but ended up abandoning this approach and developing a so-called eikonal approximation for estimating its asymptotic form. Steven Weinberg wrote a paper attacking the problem of asymptotic behaviour using similar techniques, which did not even mention the renormalization group (Weinberg, 1963). Marshall Baker, Kenneth Johnson and Raymond Wiley likewise developed their own framework for addressing the ultraviolet consistency of QED (Baker, Johnson, and Wiley, 1963, 1964, 1967). We thus see a pattern of attempts to replace, rather than to build on, the 50s renormalization group treatment of ultraviolet asymptotic behaviour.

While Bogoluibov and Shirkov had gestured at the potential broader relevance of renormalization group ideas, attempts to apply them to new problems met with limited success. Amongst Bogoluibov's school attempts were made to connect the renormalization group to work on Regge poles, a hot topic in the 60s which was a fundamental pillar of the new S-matrix programme (Arbuzov, Logunov, Tavkhelidze, 1962; Shirkov, 1963). Low and Haung (1964) criticised this work, however, arguing that Gell-Mann-Low style functional equations did not constrain the angular momentum dependence of physical quantities at fixed energy and consequently were irrelevant to the problem of Regge poles. Overall, one gets the impression that many theorists took the 1950s work to have already exhausted the fairly limited usefulness of the renormalization group concept. Wilson's work on critical phase transitions and the perturbative demonstration of the asymptotic freedom of QCD—both Nobel prize-winning achievements which hinged on renormalization group methods—dramatically boosted the renormalization groups reputation, encouraging a new generation of physicists, in a range of sub-disciplines, to learn about these techniques.

A final factor is worth considering: did interpretive confusions play a role in slowing the propagation of renormalization group ideas? A number of authors have suggested as much. Wilson and Kogut write that one of the central defects of the early renormalization group was that "the intuitive ideas were encased in a thick shell of formalism; it has required many years to peel off the shell." (Wilson and Kogut, 1974, 82). Besides Wilson, many other physicists anecdotally record finding the early renormalization group work difficult to understand. Weinberg claims to have found Bogoluibov and Shirkov's discussion impenetrable and explicitly blames their invocation of group theory:

> Bogoluibov and Shirkov seized on the point about the invariance with respect to where you renormalize the charge, and they introduced the term "renormalization group" to express this invariance. But what they were emphasizing, it seems to me, was the least important thing in the whole business. (Weinberg, 1981, 7)

What should have been emphasised instead, according to Weinberg, was the breakdown of scale invariance made possible by the need to impose renormalization conditions at a particular momentum scale, a point which he finds more explicitly prefigured in Gell-Mann and Low's discussion. This suggests the following provocative answer to the slow uptake question: Bogoluibov and Shirkov's synthesis was a mistake, which delayed the recognition of Gell-Mann and Low's insights.

We need to be careful not to read contemporary physical interpretations of the renormalization group back into the 50s here, however. While Gell-Mann and Low's discussion of quantum corrections to the Coulomb potential can, in retrospect, be viewed as containing the seed of the notion of anomalous scale breaking, neither they nor their contemporaries, articulated this idea explicitly in the 50s. As I emphasised, Gell-Mann and Low do not even have the concept of a renormalization scale. The papers of Callan and Symanzig, which became the standard references for the perturbative renormalization group in the 70s were arguably the first to make this point manifest (Callan 1970, Symanzig 1970). Furthermore, the focus on scale symmetry, and its violation, at the end of the decade was triggered by new phenomenological work on deep inelastic scattering (Bjorken, 1969), as well as theoretical work on the operator product expansion and anomalous symmetry breaking (Wilson, 1969)—see Cao (2010) for an account of these developments. It is thus more accurate, in my view, to say that the connection between the perturbative renormalization scale and the breakdown of scale invariance was made in the 70s, rather than being rediscovered.

It should also be noted that Weinberg's view of what is interpretively important and what is ancillary in renormalization group techniques is one amongst many. Shirkov gives a diametrically opposed reading, criticising Gell-Mann and Low for failing to recognise the group-theoretic nature of their argument and taking the existence of a self-similarity transformation to be at the core of the renormalization group in all of its guises (Shirkov 1999). This draws our attention back to an issue which likely did impact on the reception of the renormalization group framework in ways that are difficult to precisely determine: the existence of distinct formal structures, and physical interpretations, falling under the rubric of "the renormalization group". I want to conclude by considering how the proceeding historical investigation sheds light on the many-faced nature of the renormalization group.

5 The Many Faces of the Renormalization Group

In contemporary physics, the renormalization group is both ubiquitous and extremely varied in its manifestations. Arguably, this remains a source of confusion and misunderstandings. History gives us one way of getting a handle on this landscape of renormalization group concepts. The twin origins of the renormalization group examined here account for at least some of the heterogeneity we find in contemporary conceptions of the renormalization group. While Bogoluibov was certainly right that there is a connection between the Stueckelberg-Petermann and Gell-Mann-Low renormalization group concepts, they are not, in fact, identical and in the later history, they have occassionally come apart once more.

They are not identical because the Stueckelberg-Petermann renormalization group concept is more general. Stueckelberg and Petermann had postulated a set of transformations relating all of the different ways of fixing the arbitrary constants that appear in QFT perturbation series. Using modern terminology, we can describe their group as spanning the full space of renormalization schemes—the possible prescriptions used to assign finite values to the series coefficients. One part of fixing a scheme is the choice of a renormalization scale, and it is the freedom to vary the renormalization scale which Bogoluibov and Shirkov had focused on in their work. However, the renormalization scale does not parameterise the full space of renormalization schemes. In addition to fixing a renormalization scale, one must also fix what is sometimes called a subtraction convention—roughly speaking, a prescription that specifies how much of the series coefficient is subtracted at the scale λ. Stueckelberg and Petermann's renormalization group should include transformations between different subtraction conventions as well as different renormalization scales. Consequently, the transformation group which Bogoluibov and Shirkov had identified in Gell-mann and Low (1954) is really a sub-group (or, more correctly according to recent mathematical physics treatments, a cocycle, Dütch 2019) of the larger Stueckelberg-Petermann group.

After being largely subsumed by Bogoluibov and Shirkov's synthesis in our period of study, Stueckelberg and Petermann's concept has stepped out of their shadow in a number of contexts. In mathematical physics, a great deal of work has been done on making perturbative QFT mathematically rigorous and characterising the structures underlying it. Much of this work has its roots in Stueckelberg and Petermann's idea that the problem of ultraviolet divergences and perturbative renormalization should be understood through the lens of distribution theory. In this tradition, the existence of a transformation group linking the possible developments of the perturbative S-matrix is viewed as a fundamental result, expressing a general structural feature of perturbative QFT. Indeed, a more precise version of this statement is sometimes referred to as "the main theorem of perturbative renormalization" (Popineau and Stora, 2016). Stueckelberg and Petermann, recall, had not connected their groupe de normalisation with the scaling behaviour of a QFT and had instead postulated it as a foundationally important object in its own right. We can see this mathematical physics work as reaffirming the independent significance of the larger renormalization group Stueckelberg and Petermann had been gesturing towards.

The distinction between the Stueckelberg-Petermann and Gell-Mann-Low renormalization groups was also redrawn in the debate about renormalization scheme dependence which erupted circa 1980. While the S-matrix itself is independent of the choice of renormalization scheme, truncations of the series expansion—the approximations which are actually compared to empirical data—are not. This was never fully understood in the age of QED because the so-called on-shell renormalization scheme, which sets perturbative corrections to the photon propagator to zero for zero external momenta, allowing the renormalized charge to be identified with the charge appearing in the classical fine structure constant, was taken to be a uniquely physically motivated scheme. The scheme dependence of truncations of the perturbation series became a practical issue in QCD in part because there is no low energy classical theory we can use to define an analogous "physical" scheme. This led to a heated debate about how the renormalization scheme should be fixed in perturbative calculations. In at least some of the proposed approaches to this problem, the Stueckelberg-Petermann renormalization group, again, resurfaced as an important object in its own right. Paul Stevenson's principle of minimal sensitivity, for instance, fixes the scheme by minimising the approximants sensitivity with respect to the "full renormalization group of Stueckelberg and Petermann" (Stevenson, 1981).[21]

The future trajectory of the Gell-Mann-Low side of the 50s renormalization group is more difficult to parse, owing to the substantial innovations which took place in the later tradition. Wilson clearly saw his work as continuing the ideas about the scaling behaviour of QFTs that Gell-Mann and Low and Bogoluibov and Shirkov had developed in some sense—he continued to use the term renormalization group, after all.[22] However, the framework which Wilson developed in the 70s differ in many important respects from the renormalization group structures we have examined in this paper. Wilson's conception of the renormalization group is based on a class of coarse-graining transformations, which recursively eliminate a model's small length scale degrees of freedom, producing an effective description of its large scale behaviour. These transformations are thus not usually understood as relating equivalent formulations of the same theory, where perturbative renormalization schemes are taken to be physically equivalent notational variants.[23] They also often do not form a true transformation group, since coarse-graining transformations are not typically invertible. The class of transformations one considers in the Wilsonian approach is thus usually, strictly speaking, a semi-group. On both the formal and interpretative level, Wilson's renormalization group was something quite new (and it is crucial to emphasise that there is a large amount of formal and

[21]The scheme dependence debate is arguably also a context where authors implicitly employ different interpretations of the renormalization group, and is thus a scientific debate which the present line of investigation can potentially help clarify.

[22]When Wilson places his work in a historical context he tends to emphasise continuity with the 50s renormalization group ideas we have examined in this paper—see for instance, Wilson and Kogut (1974) and Wilson's nobel lecture (Wilson, 1984). Of course, as historians and philosophers, we should take a critical view of this rhetoric, and as I suggest briefly here there are good reasons to see Wilson's renormalization group ideas as radically novel.

[23]There is quite a bit more to say on this point as Wilsonian coarse-graining transformations sometimes are interpreted as relating physically equivalent systems—see Rosaler and Harlander (2019) for a discussion of this point in the recent philosophical literature.

interpretive diversity within the Wilsonian tradition itself which stands in need of further historical and philosophical scrutiny).

This new renormalization group did not replace the old but rather coexisted with it. The work of Callan and Symanzig in the 1970s took up and developed the implications of the group of transformations between different renormalization scales, remaining within the context of perturbative QFT. Today, this later incarnation of the Gell-Mann-Low renormalization group is much more central to the practice of perturbative QFT than it ever was in the 50s and 60s. The twin origins of the renormalization group I have examined in this paper helps us understand some of the diversity that exists in contemporary renormalization group approaches then, but the fact that new renormalization group concepts layered on top of this already imperfectly unified base is another, and ultimately perhaps more important, part of the story.

There is clearly much more historical work that remains to be done. The present study makes clear that the project of giving a more complete historical analysis of the renormalization group should not be conceived as tracing the development of a unitary scientific concept. Rather, an approach that is sensitive to the existence of many quasi-independent strands within the renormalization group tradition is needed. A useful parallel here might be with historical work on the notion of biological species, which has stressed the complex interplay between multifarious species concepts (Wilkins, 2009). There is also a need for philosophers of science to clarify the conceptual relationships between the various renormalization group notions. Arguably, mapping out this conceptual space is a necessary component of other projects which philosophical engage with renormalization group ideas. In recent years the renormalization group has been connected to debates about the foundations of QFT, scientific explanation, and emergence, to name only a few (Wallace 2011, Fraser 2020, Batterman 2002, Butterfield 2014). The present study suggests the following cautionary moral for investigations of this sort. If we ask questions, or state theses about the philosophical significance of "the renormalization group" without being sensitive to the variety of approaches that term encompasses we risk falling foul of oversimplification and terminological confusion.

Bibliography

Arbusov, B. A., Logunov, A. A., Tavkhelidze, A. N., and Faustov, R. N. (1962). "The Asymptotic Behaviour of Scattering Amplitudes and the Renormalization Group Method" Phys. Letters, 2, 150.

Batterman, R. W. (2001). *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence.* Oxford University Press.

Butterfield, J. (2014). "Reduction, Emergence, and Renormalization." The Journal of Philosophy, 111(1), 5-49.

Blum, A. S. (2017). "The State is Not Abolished, it Withers Away: How Quantum

Field Theory Became a Theory of Scattering", Stud. Hist. Philos. Mod. Phys., 60, 46-80.

Blum, A. S. (forthcoming). *The Decline and Fall of QED*, Cambridge: Cambridge University Press.

Blum. A. S, Fraser, J. D, and Miller, M. (2021). "From Infinities to Ambiguities: Ultraviolet Divergences Reconceptualised in the Causal Perturbation Theory Programme", draft manuscript.

Bogoluibov, N. N. (1955). "The Causality Condition in Quantum Field Theory" Izv. Akad. Nauk SSSR, 19, 237.

Bogoluibov, N. N. and Shirkov, (1955a). "On the Renormalization Group in Quantum Electrodynamics" Dokl. Akad. Nauk SSSR, 103, 203-206.

Bogoluibov, N. N. and Shirkov, (1955b). "Application of the Renormalization Group to Improve the Formulae of Perturbation Theory" Dokl. Akad. Nauk SSSR, 103, 391-394.

Bogoluibov, N. N. and Shirkov, (1955c). "Problems in Quantum Field Theory. I", Uspekhi Fiz. Nauk. 55, 149-214.

Bogoluibov, N. N. and Shirkov, (1955d). "Problems in Quantum Field Theory. II", Uspekhi Fiz. Nauk. 57, (1955d), 3-92.

N.N. Bogoliubov and D.V. Shirkov, (1956). "Charge Renormalization Group in Quantum Field Theory", Nuovo Cim. 3, 845-637.

Bogoluibov, N. N. and Shirkov (1959). *Introduction to the Theory of Quantum Fields* (English translation, Russian version first published in 1957)

Bjorken, J. D. (1969). "Asymptotic Sum Rules at Infinite Momentum." Phys. Rev. 179, 1547- 1553.

Bjorken, J. D., and Drell, S. D. (1965). *Relativistic Quantum Fields.* McGraw-Hill.

Dütsch, M. (2019). *From Classical Field Theory to Perturbative Quantum Field Theory.* Springer International Publishing.

Epstein, H., and Glaser, V. (1973). "The Role of Locality in Perturbation Theory. In Annales de l'IHP Physique théorique (Vol. 19, No. 3, pp. 211-295).

Fraser, J. D. (2020). "Towards a Realist View of Quantum Field Theory" in French, S. R. and Saatsi, J. (eds) *Scientific Realism and the Quantum*, Oxford University Press.

Gell-Mann, M. and Low, F. (1954). "Quantum Electrodynamics at Short Distances" Physical Review, 95, 1300.

Heisenberg, W. (1943a). "Die Beobachtbaren Grössen in der Theorie der Elementarteilchen" Zeitschrift für Physik, 120, 513e538.

Heisenberg, W. (1943b). “Die Beobachtbaren Grössen in der Theorie der Elementarteilchen. II.” Zeitschrift für Physik, 120, 673e702.

’t Hooft, G. (1999). ”When was Asymptotic Freedom Discovered? or The Rehabilitation of Quantum Field Theory”, Nucl. Phys. Proc. Suppl. 74:413–425.

Ioffe, B. L. (2002). Landau’s Theoretical Minimum, Landau’s Seminar, ITEP in the Beginning of the 1950’s. arXiv preprint hep-ph/0204295.

Källén G. (1952). “On the Definition of the Renormalization Constants in Quantum Electrodynamics” Helvica Physica Acta 25, 417

Källén G. (1953). “On the Magnitude of the Renormalization Constants in Quantum Electrodynamics” Mat-Fys Medd Dan. Vid Selesk 27, 12

Källén G. (1954). “Review: Quantum Electrodynamics at Short Distances” Zentralblatt der Mathematik

Källén, G., and Pauli, W. E. F. (1955). “On the Mathematical Structure of TD Lee’s Model of a Renormalizable Field Theory.” Dan. mat. fys. Medd., 30(CERN-55-29), 1-23.

Landau, L.D. (1955). ”On the Quantum Theory of Fields”. In Pauli, W. (ed) *Niels Bohr and the development of physics*, Pergamon, London, pp 52-69.

Landau, L. D., Abrikosov, A. A. and Khalatnikov, I. M. (1954). ”On the Elimination of Infinities in Quantum Electrodynamics.” Doklady Akad. Nauk SSSR. Vol. 95.

Landau, L. D., Abrikosov, A. A. and Khalatnikov, I. M. (1954). ”The Electron Mass in Quantum Electrodynamics.” Dokl. Akad. Nauk SSSR. Vol. 96.

Landau, L. D., Abrikosov, A. A. and Khalatnikov, I. M. (1954). ”Asymptotic Expresssion for the Green Function of Photons in Quantum Electrodynamics.” Doklady Akad. Nauk SSSR. Vol. 95.

Landau, L.D. and Pomeranchuk, I. Y. (1955). “On the Point Interactions of Quantum Electrodynamics” Dokl.Akad. Nauk SSSR, 102, 489-492.

Lehmann, H. (1954). “Über Eigenschaften Von Ausbreitungsfunktionen Und Renormierungskonstanten Quantisierter Felder.” Il Nuovo Cimento (1943-1954), 11(4), 342-357.

Petermann, A. (1953). “Divergence of Perturbation Expansion.” Physical Review, 89(5), 1160.

Popineau, G., and Stora, R. (2016). “A Pedagogical Remark on the Main Theorem of Perturbative Renormalization Theory.” Nuclear Physics B, 912, 70-78.

Rosaler, J., and Harlander, R. (2019). “Naturalness, Wilsonian Renormalization, and “Fundamental Parameters” in Quantum Field Theory.” Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 66, 118-134.

Scharf, G. (2014) *Finite Quantum Electrodynamics: The Causal Approach.* Courier Corporation.

Schwinger, J. (1949) "Quantum electrodynamics III: The Electromagentic Properties of the Electron Radiative Corrections to Scattering" Physical Review, 76, 790.

Schwinger J. (1964) *Fields and particles*, lecture, 1963 Belfer Conference (annotated typescript), in the finding guide of the Schwinger papers at UCLA.

Shirkov, D. V. (1963). "On the Structure of the Vacuum Trajectory" Phys. Lett. 3, 273.

Shirkov, D. V. (1996). "The Bogoluibov Renormalization Group" arXiv preprint hep-th/9602024

Stevenson, P. M. (1981). "Optimized Perturbation Theory." Physical Review D 23(12): 2916.

Stueckelberg, E. C. G. (1944). "An Unambiguous Method of Avoiding Divergences Difficulties in Quantum Theory" Nature, 153, 143.

Stueckelberg, E. C. G. and Rivier, D. (1950a). "A Propos des Divergences en Théorie des Champs Quantifiés" Helvetica Physica Acta, 23 (Suppl. 3), 236-239.

Stueckelberg, E. C. G., and Rivier, D. (1950b). "Causalité et Structure de la Matrice S." Helvetica Physica Acta, 23, 215-222.

Stueckelberg, E. C. G. (1951). "Relativistic Quantum Theory for Finite Time Intervals" Phys. Rev. 81, 130-133.

Stueckelberg E. C. G. and Green, T. A (1951). "Elimination des Constantes Arbitraires dans la Théorie Relativiste des Quanta" Helvetica Physica Acta, 24, 153-174.

Stueckelberg, E. C. G. and Petermann, A. (1951a). "The Normalization Group in Quantum Theory" Helvetica Physica Acta, 24, 317.

Stueckelberg, E. C. G. and Petermann, A. (1951b). "Restriction of Possible Interactions in Quantum Electrodynamics" Phys. Rev. 82(4): 548.

Stueckelberg, E. C. G. and Petermann, A. (1953). "La Normalisation des Constants dans la Théorie des Quanta" Helvetica Physica Acta, 26, 499.

Tomonaga, S. I. (1966). "Development of Quantum Electrodynamics." In Mehra, J (ed), The Physicist's Conception of Nature (pp. 404-412). Springer, Dordrecht.

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(2), 116-125.

Weinberg (1981). "Why the Renormalization Group is a Good Thing" in Guth, A. H., Huang, K., and Jaffe. R. L. (eds), *Asymptotic Realms of Physics, Essays* *in Honor of Francis E. Low* Cambridge, MA: The MIT Press.

Wilkins, J. S. (2009). *Species: A History of the Idea.*, University of California Press.

Wilson, K. G. (1961). *An Investigation of the Low Equation and the Chew-Mandelstam Equations*, PhD thesis, Caltech.

Wilson, K. G. (1969). "Nonlagrangian Models of Current Algebra." Phys. Rev., 179:1499–1512, 1969.

Wilson, K. G. (1971). "Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture." Physical review B, 4(9), 3174.

Wilson, K. G. (1971). "Renormalization Group and Critical Phenomena. II. Phase-space Cell Analysis of Critical Behavior." Physical Review B, 4(9), 3184.

Wilson, K. G., and Kogut, J. (1974). "The Renormalization Group and the ϵ Expansion." Physics reports, 12(2), 75-199.

Wilson, K. G. (1983). "The Renormalization Group and Critical Phenomena." Reviews of Modern Physics, 55(3), 583.