Representation and Realism in the Age of Effective Theories

Sébastien Rivat

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020

Abstract

Representation and Realism in the Age of Effective Theories

Sébastien Rivat

Philosophers traditionally engage with metaphysical questions at the frontiers of physics by treating theories as putatively fundamental and complete. While this interpretative strategy sits uneasily with the limited success of past theories, it breaks down with the failure of our best current theories, Quantum Field Theories (QFTs), to consistently describe the world on the smallest scales. My dissertation examines how physicists' reconceptualization of successful theories as *effective theories* affects the epistemological and semantic foundations of the interpretative practice in physics. Chapter 1 offers a detailed analysis of renormalization theory, the set of methods that underwrite physicists' construction of empirically successful QFTs. Chapter 2 demonstrates that effective theories are not merely the only candidates left for scientific realists in QFT but also worth interpreting in realist terms. Chapter 3 shows that effective theories stand as a challenge for traditional approaches to scientific representation and realism in physics. I suggest that indexing truth to physical scales is the most promising way to account for the success of effective theories in realist terms. Chapter 4 develops the referential component of this proposal by taking a detour through the problem of referential failure across theory-change. I argue that to reliably assess referential success before theory-change, we need to index reference-fixing to the limited physical contexts where a given theory is empirically reliable.

Table of Contents

List of Tables

List of Figures

Acknowledgements

There are many people whose support and encouragement were decisive for writing this dissertation and without whom I would not have enjoyed thinking about it as much as I have during the past few years.

I'm incredibly thankful to David Albert, first, for the continued care he has shown to me since my first day at Columbia and for his invaluable inputs throughout the development of this project. David helped me find my way in the labyrinth of physics, discover what I think would really matter, think about it clearly, and make it accessible to others. Above all, David taught me a distinctive and powerful way to think about physics and I can't think of any better way of having been advised. With the benefit of insight, it might be most apparent through my attempt to pay increasing attention to key toy-models, think about them in a quasi-mechanical way, and extract from them some disturbing features that might make me anxious at least for a night.

Porter Williams has also played a decisive role, as a fellow graduate student, friend, and mentor. I had been mainly taught quantum field theory in the old-fashioned way at Cambridge by physicists like Hugh Osborn and Michael Green. Porter helped me complete the transition to the effective field theory picture and get a little bit less confused about renormalization. I'm also indebted to Porter for ingraining in me a stronger sense of responsibility as a philosopher of physics and for his continued encouragement after all these years.

My work has also greatly benefited from David Wallace's generosity. Most importantly, David taught me how to be naïve about the foundations of quantum field theory in a slightly less naïve way than I originally was. I'm probably still too quick on many counts. But I can't imagine what this dissertation would have looked like without David's insights, and I'm deeply grateful for the time he took to make me think twice about central aspects of it.

I'm really thankful to Achille Varzi and Jenann Ismael for their incredible support, both intellectual and personal. Their advice and recommendations often proved crucial for navigating the ups-and-downs of graduate school and of the job market, and they both have this remarkable ability to make someone look at one's work with a new pair of eyes.

The list of those who helped me along the way is long and I will probably forget many. Jeremy Butterfield and Nazim Bouatta got me interested in the philosophy of quantum field theory in the first place back at Cambridge, and have always been encouraging and supportive since then. I also had fruitful exchanges with and received precious advice from Jonathan Bain, Robert Batterman, Alexander Blum, Adam Caulton, Hasok Chang, Justin Clarke-Doane, Frederik Denef, Alexei Grinbaum, Philip Kitcher, Tamar Lando, Michael Miller, Alberto Nicolis, Rachel Rosen, Emmanuelle Saada, Kathryn Tabb, and Nic Teh. I owe special thanks to many of my current and former fellow graduate students at Columbia too, including Sandra Baquié, Adam Blajez, Simon Brown, Alison Fernandes, Robby Finley, Thimo Heisenberg, Ye-Eun Jeong, Robbie Kubala, Joseph Lap, Umberto Mazzei, Matt Moss, Usha Nathan, Jacob Portes, Hélène Quiniou, Ignacio Ojea Quintana, Shivani Radhakrishnan, Roberto Valdovinos, Callum Watts, and Helen Zhao. I would also like to thank Maia Bernstein, Stacey Quartaro, Asha Panduranga, and Clayton Rains for their wonderful administrative work at the Department of Philosophy.

Finally, I'm deeply grateful to my family and friends outside of academia who have continuously encouraged me despite not fully understanding what I do or why I even do it. Most of all, I owe a lot to Gayatri for her continued support and unwavering care during the hard part of the writing process.

For Alain, Dominique, and Emmanuel.

Introduction

There is a long-standing belief among physicists according to which the ultimate goal of physics is to discover the final and complete theory of the physical world. This belief is perhaps sometimes too naively voiced or too easily derided. The fact remains that it resurfaces regularly in physicists' dreams, inspires much of their speculative work, and has worked as a powerful regulative ideal over the past centuries.

Philosophers of physics usually make a living by going a step further. Wilfrid Sellars once said that "the aim of philosophy [...] is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term." (1963, p. 1) Tradition has it that one of the central aims of philosophy of physics is to understand how the fundamental constituents of the world and their governing principles make everything else stick together.

Unfortunately, physicists have not yet fulfilled the promise of a final theory to any reasonable extent, if there ever was such a promise, and philosophers have no choice but to fall back on the best extant theories. Still, it has become common practice to engage with interpretative questions by treating these theories as little theories of everything. The game is then to describe the possible worlds in which these theories are exactly true, modulo some sophisticated fiddling with their structure, and find good reasons to believe that the actual world is similar enough to some of those worlds. The hope: that some of the lessons drawn from these theories, including lessons about determinism and locality for instance, will carry over to the final theory, if only approximately.

It is a central contention of this work that the new paradigm of *effective theories* in physics requires a radical revision of the foundations of this interpretative practice. The modern apparatus of effective theories was first developed in the context of particle and condensed matter physics in the 1970s-80, and has since become increasingly popular across physics thanks to its remarkable computational, heuristic, and predictive power. Yet, effective theories are not just powerful scientific devices, as their name suggests. They also have the remarkable feature (among others) of having the limits of their domain of applicability directly written in their mathematical structure. Since the consensus in physics nowadays is to think that all empirically successful theories are best understood and formulated as effective theories, philosophers' business is at risk. The best extant theories, insofar as we formulate them as effective theories, cannot, as a matter of design, be treated as little theories of everything.

This dissertation is structured around two central claims. The first concerns the epistemological foundations of the interpretative practice in philosophy of physics. I hold that effective theories furnish us with a more reliable standpoint from which to draw conclusions about the world compared to their putatively fundamental counterparts. The idea of treating physical theories as little theories of everything and learning from them in regimes where they are known with good confidence to be unreliable is already odd. The current situation with our most empirically successful and fundamental theories, Quantum Field Theories (QFTs), makes this idea untenable. In their standard formulation, the best extant QFTs become inconsistent at short distances and thus do not even provide us with meaningful physical information at these scales. A natural response would be to maintain that physics has not yet in its most fundamental quarters a sufficiently good theory to hand to philosophers. I show that this response is mistaken. Effective theories do offer a reliable avenue for understanding how physical things hang together within limited parts of the world.

The second claim concerns the semantic foundations of this interpretative practice. I show that effective theories force us to re-conceptualize the way successful theories latch onto the world if we wish to account for their success in realist terms. Scientific realists usually take the success of our best theories to provide good enough reasons to believe that some of their privileged parts are approximately true or false *simpliciter*. This traditional approach, however, does not enable interpreters to identify definite entities or structures within the limited domain of our best effective theories. I hold that the best way to avoid this issue is to take their privileged parts to be approximately true or false relative to particular physical scales. I also show that indexing reference to physical scales enables scientific realists to reliably assess whether the central terms of our best theories pick out anything real at a time they are still a live concern.

I develop these claims as follows. Chapter 1 offers a comprehensive account of renormalization theory, i.e., the set of methods developed since the 1930s to construct empirically successful QFTs. The success of these methods has puzzled philosophers for some time now. I show that this success can be explained more clearly by systematically distinguishing between the methods used to construct effective QFTs and those used to construct putatively fundamental continuum QFTs. I find that the effective approach to renormalization is more physically perspicuous and conceptually coherent, and that the issues underlying the continuum approach give additional reasons to doubt that putatively fundamental continuum QFTs are ready for metaphysical analysis.

Chapter 2 shows that effective theories are worth interpreting in realist terms in the context of QFT. As some philosophers have suggested, effective QFTs appear to be too *ad hoc* and complex to be even approximately true and thus relevant for interpretative purposes. Using the particular example of the Standard Model of particle physics, I argue that these two vices are merely apparent. Insofar as these two vices are the only reason we might have to raise doubts about the epistemic significance of effective QFTs, I take the argument to clear the path for their epistemic appraisal.

Chapter 3 has two closely related goals. The first is to show that the interpretative relevance of effective theories extends beyond the specific context of QFT. I do this by extracting common structural features shared by effective theories across physics and showing that their structure give us precise constraints for the types of entities and structures we can reliably admit in the realist inventory. The second is to show that despite their interpretative relevance in different subfields of physics, effective theories still raise a challenge for popular approaches to scientific realism and representation in specific contexts, including classical and quantum field theory. I focus on our best current QFTs, and show that the most appealing candidates for making ontological commitments in the appropriate regimes—namely, correlations, particles, and lattice fields—fail in other important respects. Continuum fields, by contrast, do not suffer from the same issues. But since their core properties extend well beyond the regime where our best effective QFTs are likely to remain trustworthy, we cannot reliably commit to these entities. I suggest that the most natural and straightforward way to escape this issue is to index truth to scales and take the descriptions of continuum fields to be approximately true or false relative to particular ranges of scales.

Chapter 4 develops the referential component of this proposal by engaging with the problem of referential failure across theory-change. The challenge here is to explain why the central terms of empirically successful theories successfully pick out entities despite apparent counter-examples in the history of science, and explain this from the standpoint of each theory, i.e., without using the terms of our best current theories as a yardstick to assess referential success. I argue that this challenge is best addressed by first indexing reference-fixing to the limited context where a theory is empirically reliable. I propose a new context-dependent theory of reference based on this idea, and show that effective theories provide a blueprint for assessing referential success along these lines.

I should emphasize that this work opens up a number of new puzzles that deserve further scrutiny. For instance, I assume that the distinction between the effective and the continuum approach to renormalization tracks two dominant traditions in the history of the renormalization program. I suspect that these two traditions find their clearest expressions in the different approaches to the renormalization group developed by Murray Gell-Mann and Francis Low in the 1950s, on the one hand, and Kenneth Wilson in the 1970s, on the other hand. Defending this point, however, would require a much more involved analysis than I can provide here. Likewise, I have not said anything about the metaphysics underlying the proposal developed in Chapters 3 and 4. I assume that interpreters make ontological commitments by taking the descriptions of our best effective theories relative to particular scales at face value. This, however, does not determine by itself what the structure of the world looks like. The most natural and straightforward picture is probably one in which the world is constituted by distinct layers overlapping in complex ways, most of which are largely insensitive to one another. Effective theories do not determine by themselves whether there is ultimately a fundamental layer or, for that matter, an overarching one. But they do give us a reliable means of understanding how unobservable things in distinct layers hang together and how these layers are sensitive to each other. We might need to give up on composition relations if we want to understand how the furniture of these different layers fit together. But there is no cause of despair here. I suspect that we should be able to give a sufficiently unified yet partial picture of the world by appealing to other types of inter-level relations.

Chapter 1: Renormalization Scrutinized

In this chapter, I propose a general framework for understanding renormalization by drawing on the distinction between effective and continuum Quantum Field Theories (QFTs), and offer a comprehensive account of perturbative renormalization on this basis. My central claim is that the effective approach to renormalization provides a more physically perspicuous, conceptually coherent, and widely applicable framework to construct perturbative QFTs than the continuum approach. I also show how a careful comparison between the two approaches: (i) helps to dispel the mystery surrounding the success of the renormalization procedure; (ii) clarifies the various notions of renormalizability; and (iii) gives reasons to temper Butterfield and Bouatta's claim that some continuum QFTs are ripe for metaphysical inquiry (Butterfield and Bouatta, 2014).

1.1 Introduction

Renormalization is one of those great success stories in physics that fly in the face of philosophers' ideals of scientific methodology. QFTs have been known to be plagued by mathematical infinities since the 1930s and it was only in the late 1940s that physicists had their first significant victory by developing appropriate renormalization techniques. It could have been hoped that they would eventually construct a realistic QFT from first principles without using these techniques; but even after seventy years, this has not been the case. Our best QFTs are still constructed by means of conceptually odd and *ad hoc* renormalization techniques. One notable example is to isolate and cancel infinite quantities by shifting the dimension of space-time by some infinitesimal amount. Another one is to simply impose some arbitrary restriction on the range of distance scales of the theory.

Among the philosophers who take the formulation of QFT most widely adopted by physicists seriously, it has become standard to appeal to the Renormalization Group (RG) theory in order to explain the unlikely success of renormalization. For instance, Huggett and Weingard (1995, sec. 2) emphasize that the RG provides the appropriate tools for identifying the class of well-defined continuum QFTs and dispels the interpretative worries related to cancellations of infinities in perturbation theory. To give another example, although with a different understanding of QFT this time, Wallace (2006, pp. 48-50; 2011, sec. 4) relies on RG-based considerations to dispel the interpretative worries related to the crude and arbitrary implementation of a physically meaningful cut-off.

Those philosophers are right to emphasize the role and the importance of the RG in contemporary physics. But there are reasons to be dissatisfied. Of central importance is the failure to appreciate the existence of conceptually distinct modern formulations of renormalization, RG included. Consider for instance Huggett and Weingard's attempt at clarifying renormalization in the case of continuum QFTs. If by 'RG' they mean the Gell-Mann & Low RG, then their account does not really dissolve the methodological worries that physicists had in the 1940s. The delicate fine-tuning of theories in the infinite cut-off limit is nothing but the old-fashioned cancellation of infinities in a different guise. On the other hand, if by 'RG' they mean the Wilsonian RG, then their account does not properly deal with continuum QFTs. At least as we traditionally understand it, the Wilsonian RG is built on the idea of integrating out high-energy degrees of freedom and restricting the applicability of the resulting theories to sufficiently large-distance scales (e.g., Weinberg, 1995, sec. 12.4; Schwartz, 2013, chap. 23).

To give another example, Cao and Schweber (1993) somewhat overstate the triumph of the modern Wilsonian renormalization programme. Many renormalization techniques conceptually akin to the approach of the late 1940s are still the "industry standard" in high energy physics, as Hollowood (2013, p. 3) felicitously puts it. These techniques include modern regularization methods such as dimensional regularization in standard QFTs and regularization by dimensional reduction in supersymmetric QFTs. More importantly perhaps, the Wilsonian RG does not fully dispel the traditional mathematical, conceptual and methodological worries associated with renormalization. With regard to methodology, for instance, one might be concerned about the infinite number of independent parameters typically required to compensate for the uncertainty associated with the exact value of a physically meaningful cut-off.

The main goal of this chapter is to offer a more accurate and systematic way of understanding the overall conceptual structure of renormalization. For this purpose, I will distinguish between the “effective" and the “continuum" approach to renormalization and show that all the important features of perturbative renormalization can be understood along this distinction. The idea is simple: current working QFTs in high energy physics are understood and formulated either as continuum QFTs or as effective QFTs, and each of these two types of QFTs is associated with a specific methodology of theory-construction—or at least, given the diversity of renormalization techniques, each of them is most conceptually consistent with a specific methodology. In the effective approach, the domain of applicability of the theory is restricted by a physically meaningful short-distance scale and the structure of the theory adjusted by including the appropriate sensitivity to the physics beyond this scale. Here, the goal is to focus on the appropriate low-energy degrees of freedom. In the continuum approach, the theory is defined across all distance scales and its structure adjusted according to the physical scale of interest. Here, the goal is to define a putatively fundamental QFT and resist the suggestion that realistic QFTs are ultimately to be understood and formulated as phenomenological theories restricted to some limited range of distance scales.

The central claim of this chapter is that the effective approach provides a more physically perspicuous, conceptually coherent, and widely applicable framework to construct perturbative QFTs than the continuum approach. I will defend this claim by showing, in detail, how the steps underlying the perturbative construction of an effective QFT are physically justified and how the resulting parts of the theory are physically meaningful, unambiguously characterized, and coherently related to one another—and this independently of the particular local QFT considered. And I will show how a careful comparison between the two approaches: (i) helps to dispel the mystery surrounding the success of the renormalization procedure discussed early on (e.g., Teller, 1988, 1989; Huggett and Weingard, 1995, 1996) but never fully dispelled in my sense, not even in the most recent literature (e.g., Butterfield and Bouatta, 2015; Crowther and Linnemann, 2017; J. D. Fraser, 2020a; 2020b); (ii) helps to clarify the various notions of renormalizability; and (iii) gives reasons to temper Butterfield and Bouatta's claim that some continuum QFTs are ripe for metaphysical inquiry (Butterfield, 2014; Butterfield and Bouatta, 2014).

The chapter is organized as follows. Section 1.2 introduces the QFT framework and the problem of ultraviolet divergences. Section 1.3 compares the effective and the continuum approach to the renormalization procedure. Section 1.4 disentangles the effective and continuum notions of perturbative renormalizability. Sections 1.5 and 1.6 briefly compare the effective and the continuum approach to the RG and clarify the scope of the continuum approach.[1] Section 1.7 examines the implications of the discussion in sections 1.3-1.6 for Butterfield and Bouatta's defense of continuum QFTs.

Three important clarifications before I begin. First, I do not think that the methodological superiority of the effective approach to renormalization offers a sufficient reason to take effective QFTs to be the correct way of understanding QFTs. It is a good step forward. But it needs to be supplemented with a careful analysis of the theoretical virtues of effective QFTs, and this goes beyond the scope of the present chapter. Second, I do not mean to claim that the distinction between the effective and the continuum approach is absolutely perfect and exhaustive. All I aim to capture is a set of salient conceptual differences that do not reduce to mere practical differences (e.g., computational simplicity and efficiency). Third, unless otherwise indicated, I will follow Butterfield (2014, pp. 30-31) and understand 'theory' in its specific sense throughout the chapter, that is to say as given by a particular action, a Lagrangian, or a Hamiltonian.

1.2 Relativistic QFT and the Problem of Ultraviolet Divergences

Relativistic QFT is the mathematical framework developed by physicists since the late 1920s to extend the tools of quantum mechanics to classical electromagnetism (and more) and to overcome the failure of quantum mechanics to account (among other phenomena) for the creation and

[1]For two recent and insightful reviews of the Wilsonian RG, see Williams (2019a) and J. D. Fraser (2020b).

annihilation of particles observed in decay experiments.

As its name suggests, a QFT describes the quantum analogue of classical fields, and the simplest way to think about a quantum field is to treat it as a continuous physical system composed of one individual quantum system at each point of space. Each individual quantum system is associated with at least one independent variable quantity (a “degree of freedom") determining the set of its possible states, and the possible states of the quantum field over space-time are obtained by combining the state spaces of these individual quantum systems together. From there, things work exactly as in quantum mechanics. A sum of states of the field (a “state superposition") also defines a possible state of the field. Each state of the field is associated with a possible configuration or “history" of the field specifying a set of values that the field can take over space-time: for instance, one real number $\phi(x)$ at each space-time point for a simple scalar field. The probability for the quantum field to be found in the configuration state $|\phi(x)\rangle$ is given by the absolute square value of the wave functional $\psi[\phi(x)]$ (assuming that we could measure the whole state of the field). And the possible energy excitation states of the field are obtained by representing the possible configuration states of the field in momentum space. One odd thing, however, is that in this picture, a “particle" corresponds to a localized pattern of energy excitations.

Quantum fields are also dynamical physical systems. They vary smoothly over space-time and interact locally at space-time points with other fields and often with themselves too. Physicists typically describe the dynamics of fields by a Lagrangian functional density $\mathcal{L}$ and the strength of interactions by coupling parameters g_i. I will take the ϕ^4-theory as my main example in what follows:

$$\mathcal{L}[\phi(x)] = -\frac{1}{2}\partial_\mu\phi(x)\partial^\mu\phi(x) - \frac{m^2}{2}\phi^2(x) - \frac{\lambda}{4!}\phi^4(x) \tag{1.1}$$

with $\phi(x)$ an arbitrary field configuration of a scalar field, m a mass parameter, and λ a quartic self-interaction coupling (using the Euclidean metric for simplicity). Of crucial importance are the action $S[\phi] = \int d^4x\mathcal{L}$ and the path integral $\mathcal{Z} = \int d[\phi(x)]e^{S[\phi]}$ which give us the different weights $e^{S[\phi]}$ associated with each possible field configuration $\phi(x)$.[2]

[2]Of course, the difficulty is that we do not yet have a mathematically rigorous definition of the path integral for

Finally, the correlations between the states of the field at n different space-time points are given by n-point correlation functions $\langle\phi(x_1)...\phi(x_n)\rangle$. Roughly speaking, these correlation functions tell us the degree to which the different "parts" of the field are sensitive to one another, i.e., here, the probability (once these functions are squared) that the field is found in a certain state at some space-time points $x_1, ..., x_k$ given its state at other space-time points $x_{k+1}, ..., x_n$ ($1 \leq k \leq n-1$). We compute empirical predictions—say, about the probability that two incoming particles decay into two outgoing particles—by calculating the absolute square value of the scattering amplitude Γ between the appropriate asymptotic particle states of the field, with Γ obtained by taking into account all the possible correlations between these states.

These are the basic tools to define and test any QFT. Unfortunately, we face two immediate problems with this "naive" schematic construction if we want to make predictions. The least severe is that realistic QFTs are highly non-linear interacting theories and therefore not exactly solvable by current mathematical means. We can still work out approximate solutions and predictions thanks to perturbation theory: provided the (dimensionless) couplings are small (e.g., $\lambda \ll 1$), scattering amplitudes can be expanded perturbatively as follows:

$$\Gamma = \lambda + \lambda^2\Gamma_2 + \lambda^3\Gamma_3 + ... \tag{1.2}$$

where each sub-amplitude $\lambda^n\Gamma_n$ represents field correlations between the incoming and outgoing particles given n possible interaction points.[3]

The most severe, the so-called problem of "ultraviolet" (UV) divergences, is that a large majority of the sub-amplitudes Γ_n actually diverge when we attempt to compute them.[4] This is clearly

realistic continuum QFTs in four dimensions, but I will ignore this problem for now.

[3]Here one might worry about two things. First, one should be wary not to interpret too quickly these perturbative terms as representing distinct real sub-processes (the so-called "virtual processes") since they might be interpreted as mere mathematical artifacts of the decomposition of Γ. Let me briefly offer one reason to resist this worry: as we will see shortly, the renormalized coupling λ is a function of an arbitrary mass scale Λ which can be interpreted as the experimental energy E at which we probe the system. Since each $\lambda^n(E)\Gamma_n(E)$ does not vary with the same rate with respect to E, we can evaluate them separately by making successive measurements at different experimental energy scales E. If this succeeds, each term receives independent empirical confirmation. Second, the perturbative series diverges in realistic cases for arbitrarily small but non-zero λ (see Helling, 2012, pp. 1-13, and Duncan, 2012, chap. 11, for more details, and Miller, 2016, for a philosophical discussion). I will ignore this problem too.

[4]I will leave aside the problem of low-energy or "infrared" (IR) divergences.

a disaster (at least at this stage) since it means that most empirical predictions in QFT are infinite. If we keep all the other assumptions of the theory in place (e.g., four space-time dimensions and standard types of fields, symmetries, and interactions), the problem naturally originates from what is known as the continuum assumption:

> **Continuum assumption**: For any extended region of space-time no matter how small, quantum fields have infinitely many interacting degrees of freedom.

In practice, the continuum assumption forces us to take into account correlations over arbitrarily short distances (or, equivalently, over arbitrarily high energies) when calculating a correlation function between any two states of the field. Consider for instance the scattering amplitude $\Gamma(p_1, .., p_4)$ in ϕ^4-theory describing the scattering event of two incoming particles decaying into two outgoing particles. Then, for example, the second-order perturbative term $\lambda^2\Gamma_2$ describes a specific set of correlations which diverge logarithmically in the high-energy domain of integration:

$$\Gamma_2 \approx \int^{\infty} d^4k/k^4 \tag{1.3}$$

with k a momentum variable. So the problem is that we have to take into account the correlations of the field over arbitrarily short distances and that the values of these correlations are small but sufficiently important once summed up to make Γ_n diverge.[5] What does it mean physically? To give a rough analogy, it is as if two distinct macroscopic parts of a table were sensitive enough to the individual particles constituting the table for the slightest movement of a particle to significantly affect on average the distance between these two parts. The sensitivity is even more dramatic in the present case. The theory is not just empirically inadequate but also inconsistent as it predicts measurement outcomes with infinite probability (i.e., here, $|\Gamma(p_1, .., p_4)|^2$ diverges).

The claim that the problem of UV divergences originates from the continuum assumption is in fact controversial, and physicists have come up with three main types of responses which I will

[5]Note that the problem does not arise in the case of non-interacting theories since there is no non-trivial correlation between distinct states in this case (i.e., $\Gamma_n = 0$ for $n \geq 1$). Note as well that in typical interacting QFTs, some contributions to the perturbative expansion are finite (e.g., box diagram integrals).

respectively call the “continuum", the “effective" and the “axiomatic" approach to the problem of UV divergences. According to the continuum approach, the problem arises because we are not working with the correct type of QFT or because we have not appropriately parametrized the QFT at hand in the first place. The hope is that the continuum assumption holds for a specific class of QFTs and that all that needs to be done is to sensibly fine-tune their parameters with the tools of renormalization theory. According to the effective approach, the problem arises because the continuum assumption is false. The solution is to impose explicit restrictions on the domain of energy scales of QFTs and adjust the sensitivity to high-energy phenomena with the tools of renormalization theory.[6] According to the axiomatic approach, the problem arises because the mathematical structure of the QFT framework is ill-defined in the first place. The solution is to develop a rigorous mathematical formulation of QFTs with explicitly stated axioms—so that, if anything goes wrong, we can at least clearly identify the origin of the problem.[7]

The crucial point is that physicists have only been able to formulate empirically successful and realistic QFTs by making extensive—if not indispensable—use of renormalization theory. It is beyond the scope of this chapter to examine the axiomatic approach, but it is worth noting here that, even after seven decades, there has not yet been any finite, exact and mathematically rigorous formulation of a realistic continuum QFT in four-dimensional space-time. If we want to understand the structure of our best current theories, a natural starting point is to look carefully at the details of renormalization.

Before delving into the details, it is instructive to start with the general idea of renormalization. Originally, renormalization was introduced as a set of techniques in high energy physics to isolate the divergent parts of scattering amplitudes and make them disappear from the final predictions

[6] Here I will ignore the specific technicalities of Effective Field Theories (EFTs) and lattice QFTs and regroup them under the category of effective QFTs together with cut-off QFTs (see, e.g., Bain, 2013; Williams, 2015; Franklin, 2018 for recent philosophical discussions about effective theories). Note, however, that lattice QFT is often understood as a specific non-perturbative regularization framework and, in this context, the goal is usually to take the continuum limit.

[7] In this context, the problem of UV divergences is usually associated with the fact that the product of distributions at a point is ill-defined (see, e.g., Steinmann, 2000, p. 73). The axiomatic criterion of rigor typically demands that the theory satisfies Wightman’s axioms (equivalently, Osterwalder-Schrader axioms) or Haag-Kastler axioms. For instance, the former includes assumptions of relativistic quantum theory, locally smeared fields, micro-causality, cyclicity, and the existence of a unique vacuum state (see Streater and Wightman, 2000). And these axioms are usually considered to be significant to the extent that they are satisfied by toy-model theories.

by absorbing them into the formal expression of the couplings of the theory. In practice, the mathematical trick works because we never directly measure the value of couplings and we can already see a similar trick at work in the simpler and more vivid case of classical electromagnetism.

Consider, for instance, the standard example of an electrostatic field produced by an infinitely long and straight wire with uniform charge density λ (per unit length), lying along the z axis of a three-dimensional Euclidean space. The measurable value of the field at some distance $r > 0$ from the wire in the xy plane orthogonal to the z direction is finite ($E \propto \lambda/r$). In contrast, the potential $V(r)$ obtained by summing up the contributions from each infinitesimal part of the wire diverges logarithmically ($V(r) \propto \lambda \int_{-\infty}^{+\infty} dz/\sqrt{z^2+r^2}$). But since we only measure differences in the values of the potential (e.g., the field $\vec{E}(x) = -\vec{\nabla}V(x)$), it makes no physical difference to subtract or add some infinite quantity in the formal expression of the potential and work with the resulting finite “renormalized" expression. One way to make this precise and well-defined is to limit ourselves to a finite portion of the wire of arbitrary length L_0 ($V_{L_0}(r) \propto \lambda \int_{-L_0/2}^{+L_0/2} dz/\sqrt{z^2+r^2}$). Subtracting the value of $V_{L_0}(L)$ for some fixed constant L to $V_{L_0}(r)$ leaves us with the finite function $-\lambda \ln(r/L)/2\pi$ and a finite residue depending on L_0 which vanishes if we take L_0 to infinity. The resulting renormalized expression of the potential is given by $V_R(r) = \lim_{L_0\to\infty} V_{L_0}(r) - V_{L_0}(L)$.

More generally, the term ‘renormalization’ designates a set of techniques used to transform the kinetic and the interacting structure of theories. On the more practical side, one finds (among others) the renormalization procedure where the main goal is to generate finite and accurate predictions. On the more theoretical side, one finds the Renormalization Group (RG) theory where the main goal is to analyze the scale-dependent structure of QFTs. As we will see in section 1.5, it is also useful to distinguish between perturbative and non-perturbative renormalization methods, even if renormalization theory is, in large part, a set of techniques specified and used in the context of perturbation theory.[8] And, finally, other areas in physics have specific renormalization techniques that I will not discuss here, such as: (i) the discretized versions of the RG in condensed

[8]For (various versions of) the Non-Perturbative Renormalization Group (NPRG) theory, see Bagnuls and Bervillier (2001), Polonyi (2003), and Delamotte (2012) for introductory reviews. Note also the existence of axiomatic renormalization methods (e.g., Scharf, 1995).

matter physics and (ii) the holographic RG in the context of gauge/gravity dualities.

1.3 Understanding the Renormalization Procedure

I argue in this section that the effective approach to renormalization offers a more physically perspicuous and conceptually coherent framework for constructing perturbative QFTs. By 'physically perspicuous' and 'conceptually coherent', I mean that the steps involved in the perturbative construction of the theory are physically justified, that the parts of the theory have a clear physical meaning, and that they are coherently related to one another. I will focus on the renormalization procedure since the main differences between the two approaches are most clearly visible at this level. The upshot is, I believe, considerable: the contrast helps dissolve the much-discussed mystery of renormalization, i.e., the issue of explaining the unlikely success of the renormalization procedure (e.g., Teller, 1988; 1989; Huggett and Weingard, 1995; 1996; J. D. Fraser, 2020a). Here, again, I should emphasize that there are many different ways to implement the renormalization procedure. I will present the steps that are most conceptually consistent with the appropriate type of perturbative QFT in each case.

1.3.1 The Effective Approach

The effective approach to the renormalization procedure is a two-step maneuver.

(i) One first *regularizes* the divergent sub-amplitudes Γ_n by introducing a limiting high-energy scale Λ_0, called the "cut-off" or "regulator". If we look at Eq. 1.3 and disregard potential trouble in the IR (i.e., $k \to 0$), $\Gamma_2(\Lambda_0) \approx \int^{\Lambda_0} d^4k/k^4$ is now a mathematically well-defined and manipulable finite quantity. But one might worry about the arbitrary choice of cut-off. A sharp cut-off separates low-energy and high-energy scales in a crude way, and we do not have enough information at this stage to decide whether an exponentially decreasing cut-off (e.g., $\int^{+\infty} d^4k \; e^{-k/\Lambda_0}/k^4$), a Gaussian cut-off (e.g., $\int^{+\infty} d^4k \; e^{-k^2/\Lambda_0^2}/k^4$) or what have you is the appropriate regulator.

(ii) The renormalization step compensates for this lack of constraint: one *renormalizes* the sub-amplitudes $\Gamma_n(\Lambda_0)$ by analyzing the relevant sensitivity to high energies and absorbing it into the couplings. The best way of implementing this idea is to include contributions to Γ_n from a specific layer of energy scales $[\Lambda, \Lambda_0]$ into a low-energy theory defined only up to Λ. Call the initial regularized theory the "bare" theory $\mathcal{L}_0(\Lambda_0)$ and its parameters the "bare" parameters λ_0 and m_0. The cut-off scale Λ_0 is the physical scale at which the theory is believed to become inapplicable and the "renormalization scale" Λ is the energy scale specifying the physics of interest, with $\Lambda \ll \Lambda_0$. In the example above, the contribution from $[\Lambda, \Lambda_0]$ is equal to

$$\begin{aligned}\lambda_0^2 \Gamma_2(\Lambda, \Lambda_0) &= \frac{3}{2}\lambda_0^2 \int_{\Lambda}^{\Lambda_0} \frac{d^4k}{(2\pi)^4} \frac{1}{(k^2+m_0^2)^2} \\ &\approx \frac{3}{16\pi^2}\lambda_0^2 \ln(\frac{\Lambda_0}{\Lambda}) \end{aligned} \tag{1.4}$$

assuming that the bare parameters are small ($\lambda_0, m_0/\Lambda \ll 1$).[9]

The essential point now is that we can simulate the effect of this high-energy contribution in the expression of the bare theory $\mathcal{L}_0(\Lambda)$ restricted to the energy scale Λ (see Fig. 1.1).[10] For that, we just need to include a new interaction term $\delta\mathcal{L}_0(\Lambda, \Lambda_0, \lambda_0) := -\frac{\lambda_{ct}}{4!}\phi^4$, called a "counter-term", with $\lambda_{ct} = -\frac{3}{16\pi^2}\lambda_0^2 \ln(\frac{\Lambda_0}{\Lambda})$. Given Eq. 1.1, this amounts to shifting the value of λ_0 to $\lambda_0 + \lambda_{ct}$, i.e., to absorbing the contributions from $[\Lambda, \Lambda_0]$ into the parameter of the theory $\mathcal{L}_0(\Lambda)$. If we keep the details explicit and restrict ourselves to the second order, the new "renormalized" scattering amplitude derived from $\mathcal{L}_0(\Lambda) + \delta\mathcal{L}_0(\Lambda, \Lambda_0, \lambda_0)$ takes the form (cf. Eq. 1.2):

$$\begin{aligned}\Gamma_R(\Lambda) &= -(\lambda_0 + \lambda_{ct}) + \frac{3}{16\pi^2}(\lambda_0 + \lambda_{ct})^2\big(\ln(\frac{\Lambda}{m_0}) - \frac{1}{2}\big) + \ldots \\ &= -\lambda_0 + \frac{3}{16\pi^2}\lambda_0^2\big(\ln(\frac{\Lambda_0}{m_0}) - \frac{1}{2}\big) + O(\lambda_0^3) \qquad . \end{aligned} \tag{1.5}$$

The renormalized *effective* theory $\mathcal{L}_R(\Lambda) := \mathcal{L}_0(\Lambda) + \delta\mathcal{L}_0(\Lambda, \Lambda_0, \lambda_0)$ defined up to Λ is obtained

[9] For simplicity, I will ignore terms in $O(1/\Lambda)$ and $O(1/\Lambda_0)$ and any field renormalization of the bare field ϕ_0 ("wavefunction renormalization factors").

[10] At this stage, one way of understanding the dependence of the Lagrangian functional density on the parameter Λ (or Λ_0) is to take it to refer to a restriction imposed on the Feynman rules used to compute scattering amplitudes in momentum space.

by defining “renormalized" parameters up to the relevant order in perturbation theory:

$$\lambda_R(\Lambda) := \lambda_0 + \lambda_{ct} = \lambda_0 - \frac{3}{16\pi^2}\lambda_0^2 \ln(\frac{\Lambda_0}{\Lambda}) \qquad . \tag{1.6}$$

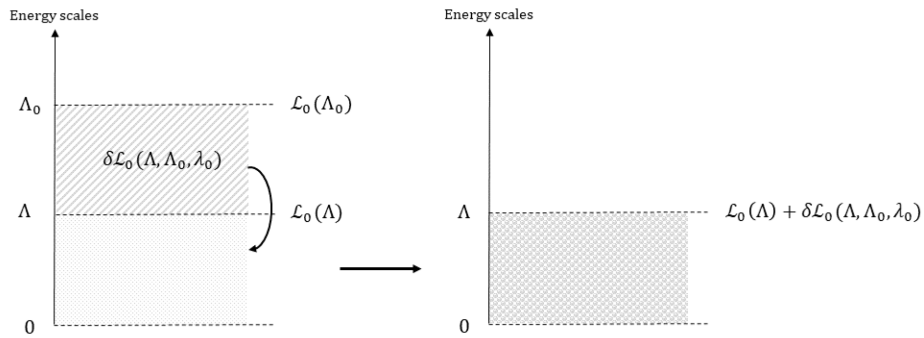

Figure 1.1: Schematic representation of the effective approach to the renormalization procedure.

This calls for two comments. First, the regularization step violates the continuum assumption only if we take the cut-off to eliminate high-energy states in the state space of the original theory. Note, however, that there is a difference between restricting the possible states of a quantum field and assuming that the quantum field is a discrete physical system composed of one individual quantum system at each point of a space-time lattice. One way to see this is to look at the following toy-model. Consider the infinite set of oscillating field configurations $\phi_a(x) = \exp(iax)$ parametrized by $a > 0$ over a one-dimensional continuous space and the corresponding infinite set of energy excitations $\tilde{\phi}_a(k) = \delta(k - \frac{a}{2\pi})$ obtained by Fourier transform. Suppose that the state space of the theory is reduced by multiplying the energy excitations by a step-function parametrized by a cut-off Λ_0:

$$\tilde{\phi}_{a,\Lambda_0}(k) = \delta(k - \frac{a}{2\pi})\theta(\Lambda_0 - k) \tag{1.7}$$

with $\theta(\Lambda_0 - k) = 1$ if $k \leq \Lambda_0$ and 0 otherwise. For $a/2\pi \leq \Lambda_0$ (i.e., for sufficiently long wavelength oscillations), the function $\theta(\Lambda_0 - k)$ does not affect the value of $\tilde{\phi}_{a,\Lambda_0}(k)$ and we obtain the original oscillating function $\phi_{a,\Lambda_0}(x) = \exp(iax)$. Otherwise, $\phi_{a,\Lambda_0}(x)$ vanishes for $a > 2\pi\Lambda_0$. So this toy-model shows that restricting the state space of the theory by a sharp high-energy cut-off implies that the possible field configurations have a minimal periodicity pattern (of wavelength $1/\Lambda_0$ here)—but it does not necessarily imply that the quantum field is discrete. To give a classical analogy, it is as if we had ignored all the possible little ripples of characteristic size smaller than $1/\Lambda_0$ in the ocean and restricted ourselves to large enough waves in order to evaluate the correlations between the oscillations of two corks floating at some macroscopic distance $1/\Lambda$ from each other.

Second, the specific counter-term $\delta\mathcal{L}_0$ leaves the theory empirically invariant, in the sense that $\mathcal{L}_0(\Lambda_0)$ and $\mathcal{L}_R(\Lambda) := \mathcal{L}_0(\Lambda) + \delta\mathcal{L}_0(\Lambda, \Lambda_0, \lambda_0)$ generate the same scattering amplitudes. The high-energy contributions to $\Gamma(\Lambda_0)$ are just parceled out among the lower-order terms of $\Gamma_R(\Lambda)$ (see Eq. 1.5). Had we chosen a different counter-term, say, $\delta\mathcal{L}_0(\Lambda, \Lambda_0, \lambda_0) + C$ with C some finite quantity, the original and modified renormalized theories would still be empirically equivalent since we only measure variations of the same renormalized scattering amplitude at different energies (e.g., $\Gamma_R(E') - \Gamma_R(E) \propto \ln(E'/E)$).[11] So the renormalization step is really a matter of reformulating the regularized theory $\mathcal{L}_0(\Lambda_0)$ in an epistemically safer way, i.e., around the scales $\Lambda \ll \Lambda_0$ where we can trust its physical content. Inversely, if we fix the value of the renormalized parameters at some specific scale, Eqs. 1.5 and 1.6 show that variations in the value of the cut-off Λ_0 can be absorbed by adjusting the value of the bare parameters, at least for a finite range of energy scales.

1.3.2 The Continuum Approach

Let us now turn to the continuum approach. It is standard in this case too to impose a regulator and split the initial regulator-dependent bare Lagrangian into a renormalized and a counter-term Lagrangian.[12] I will proceed somewhat differently by subtracting counter-terms to the physical Lagrangian. The two methods are equivalent and, most importantly, the conclusion that the continuum approach is physically ambiguous and conceptually incoherent remains the same whether we use one method or the other. The main reason for choosing the second method is that it makes the

[11] This is a particular case of "renormalization scheme dependence". I will not discuss this issue here.

[12] The method is often called "renormalized perturbation theory" because the perturbative analysis is done in terms of the physical renormalized parameters (e.g., Peskin and Schroeder, 1995, p. 326). See, e.g., Collins (1986, sec. 2.3; 2009, sec. 2) and Schwartz (2013, part III) for different ways of implementing the renormalization procedure.

conceptual differences between the effective and the continuum approach more explicit and allows us to follow more closely the original motivation of the continuum approach.

The natural starting point, then, is to think that the original theory $\mathcal{L}$ in Eq. 1.1 corresponds to the physical theory and that its parameters are fixed by experiments. Upon finding that $\mathcal{L}$ yields divergent amplitudes, we introduce a cut-off Λ_0 (regularization) and the goal of the renormalization procedure under the continuum approach is to eliminate the problematic Λ_0-dependent terms and take $\Lambda_0 \to \infty$ at the end. So, contrary to the effective approach, the physical theory of interest is the regularized theory $\mathcal{L}(\Lambda_0)$ with fixed physical parameters λ and m and not a low-energy effective theory defined only up to Λ. Likewise, the problematic Λ_0-dependent terms derived from $\mathcal{L}(\Lambda_0)$ are cancelled by adding counter-terms to *that* theory and not by adding them to some low-energy theory $\mathcal{L}_0(\Lambda)$ as defined above. This means that the counter-terms depend on λ instead of λ_0 and that the bare theory $\mathcal{L}_0(\Lambda_0) := \mathcal{L}(\Lambda_0) - \delta\mathcal{L}(\Lambda, \Lambda_0, \lambda)$, defined up to Λ_0 as well, is an intermediary construct under the continuum approach (see Fig. 1.2).[13] Finally, the parameter Λ is an arbitrary mass scale introduced to ensure that the physical expressions in the theory have a correct physical dimension, and it parametrizes the particular choice of counter-term: e.g., $\delta\mathcal{L}(\Lambda, \Lambda_0, \lambda) = \delta\mathcal{L}(\Lambda', \Lambda_0, \lambda) + C$ for C some finite quantity and Λ' a new definition of the arbitrary mass scale. I will label the renormalization scale μ instead of Λ in the continuum approach in order to keep track of the difference of interpretation.

Once all the divergent terms have been removed up to some order n in the original expression of Γ, we can stop the renormalization procedure and safely take the limit $\Lambda_0 \to \infty$ in the renormalized expression of Γ. By assumption, the value of the physical parameters λ and m is fixed (e.g., to their experimental value measured at some energy scale E). So, by taking the limit $\Lambda_0 \to \infty$, we are required to take the limit of the bare parameters too. In our example, λ_0 diverges:

$$\lim_{\Lambda_0 \to \infty} \lambda_0 = \lim_{\Lambda_0 \to \infty} \big(\lambda + \frac{3}{16\pi^2}\lambda^2 \ln(\frac{\Lambda_0}{\mu})\big) = +\infty \tag{1.8}$$

[13] In renormalized perturbation theory, the bare Lagrangian corresponds to the initial Lagrangian with the "wrong" parameters, i.e., with the parameters that we split into a finite and an infinite part in order to cancel divergences.

In principle, the original scattering amplitude Γ can be made finite at any order by repeating the procedure. And if we know the experimental values of λ and m at the scale E, we can directly compute the quantum corrections obtained at some higher energy scale E'.

Complications arise once we realize that the formal expression of the finite renormalized scattering amplitude Γ_R still depends on the arbitrary value of the mass scale μ. Since this amplitude is supposed to be a physical amplitude, we have to assume that its formal expression does not depend on some arbitrary choice of μ. This has interesting consequences.[14] First, the value of the bare parameters does not depend on μ while the value of the original parameters depends on μ, as it can be easily seen from the expression of Γ_R:

$$\begin{aligned}\Gamma_R &= -\lambda + \frac{3}{16\pi^2}\lambda^2\big(\ln(\frac{\mu}{m}) - \frac{1}{2}\big) + ... \\ &= -\lambda_0 + \frac{3}{16\pi^2}\lambda_0^2\big(\ln(\frac{\Lambda_0}{m}) - \frac{1}{2}\big) + O(\lambda_0^3)\end{aligned} \tag{1.9}$$

This means that the original theory is a particular case of a more general *renormalized* theory $\mathcal{L}_R(\mu)$, defined in terms of renormalized parameters $\lambda(\mu)$ and $m(\mu)$. Second, in the absence of experimental measurement, we can give an explicit perturbative definition of the renormalized parameters by redefining them order by order in terms of the “fixed" (i.e., μ-independent) bare parameters (i.e., $\lambda_0 = \lambda(\mu) + O(\lambda^2(\mu)) \longrightarrow \lambda(\mu) = \lambda_0 - O(\lambda^2(\mu))$). As a result, the general renormalized theory is defined perturbatively by fine-tuning the expression of the Λ_0-dependent bare theory with the help of counter-terms:

$$\mathcal{L}_R(\mu) := \lim_{\Lambda_0 \to \infty} \mathcal{L}_0(\Lambda_0) + \delta\mathcal{L}(\mu, \Lambda_0, \lambda_0) \tag{1.10}$$

Note that the correction $\delta\mathcal{L}$ takes the form of the original counter-term as defined in the effective approach in this simple case.

Let me make one crucial comment. The finite renormalized amplitude $\Gamma_R := \lim_{\Lambda_0 \to \infty}(\Gamma - \Gamma_{ct})$

[14] Note that if we choose to fix the value of μ at some energy scale E, the resulting amplitude still depends on the arbitrary choice of counter-term, which we can parametrize by introducing some μ'.

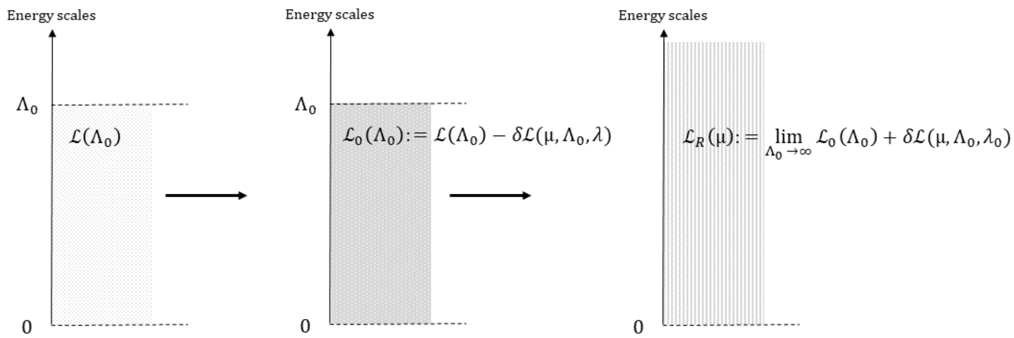

Figure 1.2: Schematic representation of the continuum approach to the renormalization procedure.

obtained by subtracting the appropriate Λ_0-dependent terms Γ_{ct} in the original expansion Γ is derived from the expression of the bare Lagrangian $\mathcal{L}_0(\Lambda_0)$.[15] Both the original and the general renormalized theory yield divergent amplitudes $\Gamma(\Lambda_0, \lambda)$ and $\Gamma(\Lambda_0, \lambda(\mu))$ in the limit $\Lambda_0 \to \infty$ if we do not restrict the state space of the theory. Similarly, in the method where the bare Lagrangian is split into a renormalized and a counter-term Lagrangian, the finite renormalized amplitude is derived from the initial bare Lagrangian with the wrong parameters, and the physical renormalized theory yields divergent amplitudes in the limit $\Lambda_0 \to \infty$.

1.3.3 Comparing the Effective and the Continuum Approach

Let me now explain why the effective approach offers a more physically perspicuous and conceptually coherent formulation of renormalization. To begin with, in the somewhat naive approach taken so far, the bare theory of most QFTs makes no physical sense under the continuum approach. The reason is that most QFTs, including Quantum Chromodynamics (QCD) and Quantum Electrodynamics (QED), are plagued with UV divergences and these divergences are cancelled by choosing the bare couplings to diverge exactly in the same way. Even in QCD, the naive perturbative expression of the bare coupling parameter between quarks and gluons takes the form of a series in the physical coupling parameter with increasingly divergent Λ_0-dependent terms at each order (see, e.g., Collins, 2011, sec. 3.3). So if we take the limit $\Lambda_0 \to \infty$ at this level, the

[15] To see this, note that the expression $\Gamma - \Gamma_{ct}$ at second order in λ is obtained from Γ_0 by expressing λ_0 in terms of λ at each order.

bare coupling diverges and the resulting bare Lagrangian is ill-defined (e.g., as in our example, $\lim_{\Lambda_0 \to \infty} \lambda_0 = \pm\infty$).[16] This means that the bare theory used in the renormalization procedure under the continuum approach is nothing more than a physically meaningless intermediary mathematical tool to generate finite renormalized scattering amplitudes. Therefore, we cannot explain the empirical success of the renormalized amplitudes by pointing at their successful derivation by means of some more general law and additional conditions since the bare Lagrangian, i.e., what plays the role of the general law here, has no physical meaning. Appealing to the renormalized Lagrangian does not help either since it generates divergent amplitudes.

A somewhat less naive approach is to realize that the perturbative expression of the bare parameters does not depend on the renormalization scale μ. If we take $\mu = \Lambda_0$ before taking the infinite cut-off limit, the bare parameters are equal to the renormalized parameters defined at the cut-off scale Λ_0. As we will see in sections 1.5 and 1.6, the appropriate perturbative expression of the renormalized parameters is obtained by means of RG methods. And we will see that even from the perspective of the RG, it turns out that there are still many QFTs for which the bare parameters diverge. There are even QFTs for which it is impossible to take the infinite cut-off limit without affecting the expression of the renormalized parameters (because of the existence of a so-called "Landau pole", see section 1.6). At any rate, all of these cases leave us in exactly the same situation as above. But perhaps the continuum approach offers a physically perspicuous and conceptually coherent formulation of renormalization only in well-behaved cases (i.e., $0 \leq \lim_{\Lambda_0 \to \infty} \lambda_0 \ll 1$). For instance, when we take $\mu = \Lambda_0$, the expression of the bare coupling in QCD converges to zero in the infinite cut-off limit and so the bare theory does not seem to be plagued with the same issues as the bare theory in the ϕ^4-theory example.

Still, it turns out that the continuum approach faces important interpretative difficulties and suffers from severe conceptual ambiguities even in well-behaved cases. First, the renormalized theory yields divergent perturbative amplitudes if we do not restrict the state space of the theory. This should at the very least refrain us from taking this theory at face value too quickly (see

[16] More generally, all the cases where the perturbative assumption $\lambda_0 \ll 1$ breaks down are pathological.

section 1.7). Second, the conceptual status of counter-terms is ambiguous under the continuum approach, and this is independent of the value of the bare parameters. Recall that whether we add the counter-terms to the original theory or obtain them by splitting the initial bare parameters into two pieces, the main role of the counter-terms is to make the original amplitude finite. We might attempt to clarify their conceptual status in two different ways. (i) Counter-terms correspond to surplus component parts of the bare theory which cancel out with other divergent parts of the bare theory when we calculate amplitudes. That is, by adding counter-terms to the physical theory, we simply re-arrange the structure of the bare theory in such a way that its superfluous divergent parts cancel each other. (ii) Counter-terms correspond to scaling factors relating the parameters of the bare and physical theories. That is, by adding counter-terms to the physical theory, we simply reparametrize the original parameters in such a way that the resulting theory, i.e., the bare theory, yields finite predictions. In both cases, however, the counter-terms cancel out precisely because we choose the component parts of the bare theory or the scaling factors of the physical theory in such a way that they cancel the divergent parts of the original amplitudes. That is, in both cases, it seems difficult to escape the conclusion that counter-terms are introduced just for the purpose of canceling divergences, which makes the whole renormalization procedure look *ad-hoc*. Moreover, it is hard to see how one could interpret the counter-terms in physical terms, including those relating the original field variable and the bare field variable (which I ignored for simplicity), and therefore to make sense of the relationship between the bare and the renormalized theory. The counter-terms are, as it were, intrinsically meaningless formal tools to derive finite predictions.

The contrast with the effective approach is striking. First, recall that on this approach, we start with the assumption that the bare theory breaks down at some physically meaningful scale Λ_0. The structure of the theory may give us very good internal reasons to believe that it becomes inconsistent at some point beyond this scale, or we may have very good external reasons to believe that the theory starts to make empirically inaccurate predictions around this scale. Either way, we take the domain of applicability of the theory to be restricted by some limited range of energy scales. On this assumption, $\mathcal{L}_0(\Lambda_0)$ is naturally interpreted as the most fundamental formulation of the theory, i.e., the theory defined up to the scale Λ_0 where it is supposed to break down. The renormalized theory $\mathcal{L}_R$ is naturally interpreted as a more physically reliable low-energy effective version of the bare theory. If we take $\Lambda = \Lambda_0$, the physical renormalized theory $\mathcal{L}_R(\Lambda)$ simply corresponds to the bare theory $\mathcal{L}_0(\Lambda_0)$ for some appropriate choice of counter-terms. And insofar as Λ_0 is kept fixed, both theories yield finite predictions, are mathematically well-defined (at least according to physicists' standards), and even yield exactly the same scattering amplitudes if we choose the counter-terms appropriately. Second, the effective approach offers a physically salient interpretation of counter-terms: whether we fix the parameters of the bare theory or those of the renormalized theory, the counter-terms are naturally interpreted as standing for high-energy effects described by the bare theory. Moreover, the introduction of counter-terms is physically justified on the grounds that the low-energy scales are not completely insensitive to the high-energy ones.

All of this should help to clarify the mystery surrounding the renormalization procedure discussed in the literature (e.g., Teller, 1988; 1989; Huggett and Weingard, 1995; 1996; J. D. Fraser, 2020a). The mystery, if anything, is a mystery about the continuum approach: it arises because the meaning and the status of the bare theory, the renormalized theory, and the counter-terms are ambiguous, and because the method used for deriving the renormalized theory and the finite renormalized scattering amplitudes is physically unjustified. By contrast, the effective approach relies on well-specified physical concepts and offers a clear physical picture of inter-scale sensitivity. The effective approach also offers a better rationale for each step of the renormalization procedure: while there are good reasons to expect a physical theory to break down at short distances (regularization step), it does not mean that it automatically fails to provide physically relevant and empirically accurate descriptions at larger distances if the relevant sensitivity to short distances is taken into account (renormalization step).

Now, the mystery surrounding the continuum approach is not as mysterious as it might seem, at least in this simple case.[17] It is a standard principle in physics that physical expressions must have the same physical dimension upon mathematical transformation for them to remain physi-

[17]More generally, the following explanation works for QFTs displaying logarithmic divergences (e.g., QED, QCD).

cally meaningful. This principle requires us to introduce the new arbitrary parameter μ with the introduction of the regulator Λ_0 (e.g., to use $\lambda^2 \ln(\Lambda_0/\mu)$ instead of $\lambda^2 \ln(\Lambda_0)$ as a counter-term). In this specific example, this principle also ensures that the arbitrary parameter μ captures exactly the sensitivity to high energies as parametrized by the regulator Λ_0, as it can be seen from the expression of the counter-term. The continuum approach therefore successfully offers a measure of the sensitivity of the low-energy physics to the high-energy physics, and this is in fact all that is needed to explain the empirical success of the theory. Had we chosen a different counter-term, say, $\lambda^2 \ln(\Lambda_0/\mu) + C$, with C some finite quantity, the same sensitivity would be captured by some appropriate redefinition of μ. Hence, even if the continuum approach offers a highly formalistic and instrumental framework, it remains at least possible to identify the reasons for its empirical success. Needless to say, the effective approach offers a more physically perspicuous and conceptually clear explanation.

Let me conclude this section by responding to two potential concerns. First, taking the limit $\Lambda_0 \to \infty$ under the effective approach does not turn the situation around. Agreed, there is nothing problematic if the goal is to probe the mathematical structure of the theory, or if we add by hand a high-energy cut-off afterwards. But, strictly speaking, taking the limit $\Lambda_0 \to \infty$ is conceptually incoherent since the introduction of the cut-off Λ_0 is justified on the grounds that it marks the physical scale at which the theory is supposed to break down. Another option is that, by taking $\Lambda_0 \to \infty$, we are actually making the approximation that the low-energy physics is largely insensitive to the high-energy physics beyond Λ_0. But in this case, it is implicitly assumed that the theory is restricted to low energies and that it should not be used to make predictions at arbitrarily high energies.

Second, the distinction between the effective and the continuum approach does not crucially depend on the specific regularization method we choose and on the specific way we subtract divergences or absorb the appropriate sensitivity to high energies (although I will emphasize in section 1.5 that each approach is most conceptually consistent with its own specific type of regularization and renormalization method). In particular, the distinction between the effective and the continuum approach does not reduce to the distinction that Georgi (1992, 1993) and Bain (2013) draw between Wilsonian and continuum EFTs. This distinction is mainly based on whether the split between the low-energy and high-energy physics depends on the mass parameter of the theory (see Georgi, 1993, sec. 1.2; Bain, 2013, sec. 4). And continuum EFTs are called "continuum" because the most famous mass-independent regularization method, namely, dimensional regularization, does not eliminate high-energy states in the state space of the theory. This, however, does not mean that continuum EFTs are meant to be used to make predictions across all energy scales. In particular, they are restricted by the energy scale characterizing the matching conditions.

1.4 (Perturbative) Renormalizability Yes... But Which One?

We have seen that the continuum approach to the renormalization procedure offers a highly formalistic and instrumental perturbative framework to derive consistent and empirically relevant predictions. It turns out that the situation is even worse for the continuum approach since the procedure only works at every order in perturbation theory for the restricted class of "perturbatively renormalizable" QFTs. After distinguishing between two distinct notions, one for the continuum approach and the other for the effective approach, I argue in this section that the continuum approach is all the less attractive as it fails to apply to a large number of successful and physically significant theories. We will see in sections 1.5 and 1.6 that the RG does not substantially affect this claim.

First consider the continuum approach. Here the notion of perturbative renormalizability is best introduced by noting that the ϕ^4-theory example used so far is extremely fortunate. All the divergent terms depending on positive powers of Λ_0 or $\log(\Lambda_0)$ that appear in the perturbative expansion of $\Gamma(p_1, .., p_4)$ can be absorbed by introducing counter-terms that depend only on the coupling λ. All the finite terms depending on positive powers of $1/\Lambda_0$ vanish as we take $\Lambda_0 \to \infty$. More generally, all the divergences that appear in any sub-amplitude can be cancelled by using only λ and m in the ϕ^4-theory. There are many theories, however, for which infinitely many new couplings need to be introduced—the 4-Fermi theory is one such example (see, e.g., Schwartz, 2013, chap. 22)—and the difference between this example and the ϕ^4-theory can be captured as follows:

> A theory is perturbatively renormalizable iff we only need to introduce a finite number of independent couplings in order to eliminate divergences and define $\mathcal{L}_R(\mu)$ at any order in perturbation theory in the limit $\Lambda_0 \to \infty$.
>
> A theory is perturbatively non-renormalizable iff we need to introduce an infinite number of independent couplings.

This characterization is of course somewhat superficial. According to Dyson's criterion, what makes a theory perturbatively non-renormalizable is that it contains *at least* one "non-renormalizable" individual interaction term, i.e., an interaction term parametrized by a coupling g_i with strictly negative mass dimension Δ_i.[18] These types of interactions generate an infinite number of sub-amplitudes with an increasing degree of divergence, and each of the resulting types of divergent quantities usually requires the introduction of a new counter-term. In contrast, the so-called "renormalizable" ($\Delta_i = 0$) and "super-renormalizable" ($\Delta_i > 0$) interaction terms generate only a finite number of different types of divergences.[19] Having said that, perturbative non-renormalizability is not a dead end. In general, both perturbatively renormalizable and non-renormalizable theories are "renormalizable" in the sense that the structure of the theory is such that it is possible to construct a counter-term to cancel any type of divergence at any order in perturbation theory. I will call this notion renormalizability$_{RT}$ to avoid confusion as it is sometimes referred to as the

[18]The mass dimension Δ of a physical quantity is the power of that quantity expressed in terms of some energy variable (i.e., energy$^{\Delta}$) with natural units $c = \hbar = 1$.

[19]The longer explanation is based on the so-called "power-counting" argument (e.g., Weinberg, 1995, sec. 12.1). A divergent integral $I = \int^{\infty} dk k^{D-1}$ is characterized by the value of its superficial degree of divergence D (the integral diverges in the UV if $D \geq 0$) and D can be expressed in terms of the mass dimensions Δ_i of the interactions involved in the scattering process described by I: schematically, $D =$ positive number $- \sum_i n_i \Delta_i$, with n_i the number of times we need to use the interaction i to define the integral. Then, if there is at least one non-renormalizable interaction in the theory ($\Delta_i < 0$), it is possible to find infinitely many different types of divergent integrals ($D \geq 0$) by considering more and more complex sub-amplitudes at higher orders in the perturbative expansion. By contrast, D has a positive upper bound for (super-) renormalizable theories, i.e., there is only a finite number of different types of divergent integrals. Note, however, that the superficial degree of divergence is not always reliable: there are cases where $D < 0$ and the integral diverges (notably because of the so-called "sub-divergence" problem), and cases where $D \geq 0$ and the integral is finite (usually the divergence cancels because of symmetry constraints). Perturbatively renormalizable theories are sometimes called "renormalizable in the power-counting sense" or "renormalizable in Dyson's sense".

"Renormalization Theorem" (e.g., Osborn, 2016, sec. 4.3).[20] We can even take the limit $\Lambda_0 \to \infty$ in a number of expressions obtained from perturbatively non-renormalizable theories at each finite order in perturbation theory if the theory is not too exotic (i.e., if $\lim_{\Lambda_0 \to \infty} g_0$ formally exists for each bare coupling g_0 given some fixed finite order in perturbation theory).

At first sight, it seems that the distinction between perturbatively renormalizable and non-renormalizable theories captures the amount of work needed in order to renormalize a theory—and the amount is of course infinite if we want to define the parameters of a non-renormalizable theory at every order in perturbation theory. In fact, the notion of perturbative renormalizability provides a much stronger criterion of theory-selection under the continuum approach. If the perturbative expression of a non-renormalizable theory is defined by introducing an infinite number of new parameters, it means that quantum corrections to scattering amplitudes depend on the specification of an infinite number of constants and that we therefore need an infinite number of experiments in order to fix their value. Since this is impossible in practice, the perturbative formulation of non-renormalizable theories obtained by applying the renormalization procedure at every order in perturbation theory turns out to be empirically useless. We should therefore restrict the class of empirically relevant theories to perturbatively renormalizable theories under the continuum approach.

So far, the analysis only applies to the continuum approach and one might wonder whether there is any equivalent notion of perturbative renormalizability under the effective approach and, if so, whether it plays the same role. Let me suggest the following notion of perturbative renormalizability$_E$, to be distinguished from the traditional notion and the notion of renormalizability$_{RT}$:

> A theory is perturbatively renormalizable$_E$ iff for any $p \in \mathbb{Z}$, all the possible contributions to predictions up to order $O((\Lambda/\Lambda_0)^p)$ obtained from $\mathcal{L}_0(\Lambda_0)$ can be absorbed in $\mathcal{L}_R(\Lambda)$ by introducing only a finite number of new parameters. (*mutatis mutandis* for perturbatively non-renormalizable$_E$.)

[20] In contrast, the term 'non-renormalization theorem' usually refers to a specific result to the effect that a parameter or an interaction term does not need to be renormalized at all at any order in perturbation theory, as it is common in supersymmetric QFTs (see, e.g., Weinberg, 2000, sec. 27.6). Of course, in practice, the interesting question is whether a theory is renormalizable$_{RT}$ *given* a set of constraints imposed on the construction of counter-terms (e.g., that they leave the resulting Lagrangian invariant under the action of a given symmetry group).

The basic idea is the following: a theory is perturbatively renormalizable$_E$ if we can always simulate high-energy effects up to a specific accuracy ϵ with only a finite number of couplings, and perturbatively non-renormalizable$_E$ if we cannot. It is not entirely clear what perturbatively non-renormalizable$_E$ theories would look like. Presumably, these types of theories would have to include exotic interaction terms such that the contributions of these terms vary too rapidly between Λ_0 and Λ to be approximated by the contributions of a finite number of independent polynomial interaction terms in the field variables and their derivatives given some accuracy ϵ. For instance, we can imagine a theory with an exotic non-local interaction term including some non-analytic function $F(\phi(x), \phi(y))$ of field variables specified at distinct space-time points x and y such that the contributions of F vary too abruptly between Λ_0 and Λ to be approximated by the contributions of a finite number of independent polynomial interaction terms.

Be that as it may, the notion of perturbative renormalizability$_E$ is much less constraining than the traditional notion of perturbative renormalizability. Perturbative renormalizability$_E$ is satisfied if the interaction terms of the theory are local polynomials in the field variables and their derivatives and if the theory has a finite number of independent interaction terms with the same dimension Δ_i. Most crucially, the notion of perturbative renormalizability$_E$ does not prevent the theory from including non-renormalizable interaction terms. Quite the contrary: under the effective approach, we often need to introduce non-renormalizable terms into the effective theory if we want to absorb contributions in $O((\Lambda/\Lambda_0)^p)$ $(p > 0)$ obtained, say, from the renormalizable interaction terms of the bare theory.[21] There is no specific reason to worry about these contributions in the continuum approach since they cancel out in the limit $\Lambda_0 \to \infty$. But to the extent that we keep the cut-off fixed, we usually need to include non-renormalizable terms in the effective theory if we want to maximize the match between the effective and the bare theory.

As a corollary, if we keep the cut-off fixed, perturbatively non-renormalizable theories remain perfectly predictive and empirically relevant. Typically, it is sufficient to consider interaction terms

[21] Lepage (1989, sec. 2.3) provides a concise explanation of this pattern. If we consider again the superficial degree of divergence of integrals (see footnote 19), it is possible for any renormalizable interaction to generate infinitely different types of finite integrals with negative superficial degrees of divergences, i.e., with contributions in $O((\Lambda/\Lambda_0)^p)$ $(p > 0)$. For some examples, see Schwartz (2013, chap. 21).

with dimension larger or equal to $\Delta_\epsilon = -\ln(\epsilon)/\ln(\Lambda/\Lambda_0)$ in order to make predictions at the energy scale Λ with accuracy ϵ (e.g., Georgi, 1993, p. 214). The total number of interaction terms with $\Delta \geq \Delta_\epsilon$ is finite in standard QFTs and we can increase the empirical accuracy of the theory by adding non-renormalizable interaction terms with $\Delta < \Delta_\epsilon$ (keeping in mind that the mass dimension of non-renormalizable interaction terms is negative). In general, the most empirically successful and physically informative version of an effective theory (the so-called "Wilsonian" effective Lagrangian) includes all the possible interaction terms compatible with the assumptions of the theory—in particular, its symmetries.[22] To give an example, the effective Lagrangian $\mathcal{L}_W$ generalizing the ϕ^4-theory takes the following form:

$$\begin{aligned}\mathcal{L}_W = -\frac{1}{2}(\partial\phi)^2 - \frac{m^2}{2}\phi^2(x) - \frac{\lambda}{4!}\phi^4(x) - \sum_{n\geq 3} g_{2n}\phi^{2n} \\ - \sum_{n\geq 2} g'_{2n}(\partial\phi)^{2n} - \sum_{n,m\geq 1} g''_{2n,2m}(\partial\phi)^{2n}\phi^{2m}\end{aligned} \tag{1.11}$$

Non-renormalizable interaction terms are those associated with the couplings g_{2n} with $n \geq 3$, g'_{2n} with $n \geq 2$, and $g''_{2n,2m}$ with $n, m \geq 1$ in this example.

Perturbatively non-renormalizable theories have been much appraised in the recent physics and philosophical literature (e.g., Lepage, 1989; Cao and Schweber, 1993; Butterfield and Bouatta, 2015; Williams, 2019a). I do not have much to add here, except the following four important points. First, for any approach, the restriction to a finite number of independent couplings is necessary in practice if we want to make empirical predictions. Second, the effective approach provides a clear physical justification for the introduction of an infinite number of additional non-renormalizable interaction terms: they capture the full sensitivity of low-energy physics to high energies, even the most insignificant parts of it. Third, perturbative renormalizability *remains* a decisive criterion of theory-selection for the perturbative formulation of *continuum* theories insofar as it is possible to define (at least formally) perturbatively non-renormalizable theories at every order

[22]A complication comes from anomalies: i.e., the renormalization procedure might require new terms which explicitly break the symmetries of the theory. This is called "anomalous" or "quantum" symmetry breaking, but I will ignore this problem here.

in perturbation theory in the $\Lambda_0 \to \infty$ limit. Fourth, the notion of perturbative renormalizability$_E$ under the effective approach offers a highly inclusive criterion of theory-selection and, as far as I can tell, all the traditional perturbatively renormalizable and non-renormalizable QFTs are perturbatively renormalizable$_E$. In a way, perturbative renormalizability$_E$ is as non-constraining as the notion of renormalizability$_{RT}$ discussed above (but less general, though).

Now, it is a matter of fact that perturbatively non-renormalizable theories have proven to be extremely useful in deriving successful empirical predictions and describing physically relevant patterns at different energy scales, from low-energy effective phenomenological models to extensions of QFTs beyond the Standard Model. This success, however, requires us to explicitly restrict the domain of applicability of these theories by means of some finite cut-off. For if we attempt to define the perturbative formulation of these theories across all scales and derive exact predictions without making any approximation, we will find that these theories lose their predictive power and empirical relevance. Of course, if we have empirical inputs and restrict ourselves to some finite order in perturbation theory, we may take the limit $\Lambda_0 \to \infty$ at this order and use the perturbatively non-renormalizable theory to make predictions. For instance, if we know the value of the Fermi constant $G_F \sim 10^{-5}$ GeV^{-2}, we can use the 4-Fermi theory to make tree-level predictions. However, if we endorse the continuum approach and intend to renormalize theories at every order in perturbation theory, we will be forced to rule out a large class of empirically and physically relevant theories. And so insofar as we want to praise a framework for constructing perturbative QFTs that proves to be (sufficiently) universal, the effective approach looks more attractive than the continuum approach.

1.5 The Renormalization Group Theory

What has been at the center stage of the renormalization procedure so far is the attempt to address the problem of UV divergences:

(1) How can finite and accurate predictions be obtained if the original theory is inconsistent?

We have seen that in both the effective and the continuum approach, the introduction of an arbitrary mass scale Λ (or μ) is forced upon us if we want to derive the expression of renormalized quantities. The genius of the physicists who developed the Renormalization Group (RG) theory was to use this seemingly idle and arbitrary parameter as a lever to address the (new) questions:

(2) What is the scaling behavior of the theory?

(3) Does the theory make consistent predictions in the continuum limit?

The goal of this section is to show how the RG theory clarifies the notion of renormalizability and therefore complicates the argument of section 1.4. Of crucial importance is the possibility that a theory both includes non-renormalizable interaction terms and makes consistent predictions in the continuum limit. At the same time, some perturbatively renormalizable theories such as the Standard Model of particle physics are likely to make inconsistent predictions at very high energies. This suggests that the scope of the continuum approach might not be as restricted as initially thought—and yet still be restricted in important ways.

1.5.1 The Effective and the Continuum RG

What, exactly, is the RG? Strictly speaking, the RG refers to the structure of invariance of theories under rescaling by the renormalization scale Λ (or μ). It is helpful, though, to distinguish between three types of RG equations. First, at the level of theories, the RG describes how the path integral, the action, and the Lagrangian transform under rescaling. In a way, the renormalization procedure already gives us a rudimentary RG transformation: e.g., in the effective approach, $\mathcal{L}_0(\Lambda_0) \to \mathcal{L}_R(\Lambda) = \mathcal{L}_0(\Lambda)+\delta\mathcal{L}_0(\Lambda, \Lambda_0)$ for $\Lambda_0 \to \Lambda$. Second, at the level of scattering amplitudes and correlation functions, the RG describes the specific trade-off between the kinetic and interacting parts of the theory required for the scattering amplitudes to remain invariant under rescaling. The so-called “Callan-Symanzik" equation for a N-particle amplitude with one renormalized coupling g is given by:

$$(\Lambda\frac{\partial}{\partial\Lambda} + \beta\frac{\partial}{\partial g} + N\gamma_\phi)\Gamma_R(p_1, ..., p_N; g(\Lambda)) = 0 \tag{1.12}$$

where $\beta(g) = \Lambda \frac{\partial g}{\partial \Lambda}$ is the "beta-function" of the coupling g and γ_ϕ is the "anomalous dimension" of the field. Eq. 1.12 describes how much we need to shift the value of the coupling ($\beta \frac{\partial}{\partial g}$) and the size of the field configurations ($N\gamma_\phi$) in order to absorb an infinitesimal rescaling ($\Lambda \frac{\partial}{\partial \Lambda}$) and leave the amplitude Γ_R invariant. Third, at the level of couplings and local operators, the RG describes how the strength of an interaction varies across scales in accordance with the sign of its beta-function. For instance, the quartic interaction in the ϕ^4-theory becomes increasingly strong at high energies:

$$\Lambda \frac{\partial \lambda_R}{\partial \Lambda} = \beta(\lambda_R) = \frac{3}{16\pi^2}\lambda_R^2 + \mathrm{O}(\lambda_R^3) \tag{1.13}$$

Note, however, that this perturbative RG equation remains only valid for $\lambda_R \ll 1$.

The effective (or Wilsonian) RG and the continuum (or Gell-Mann & Low) RG have a relatively similar formal structure overall. But again, there are significant conceptual differences between the two.[23] Most crucially, the effective renormalized theory is obtained by integrating out high-energy field configurations in the path integral, while the continuum renormalized theory is obtained by fine-tuning the expression of the bare theory. Schematically,

$$\begin{aligned} &\text{Effective theory:} \qquad && \int d[\phi_{<\Lambda}] e^{S_{\text{eff}}(\Lambda,\Lambda_0)} = \int d[\phi_{<\Lambda_0}] e^{S_0(\Lambda_0)} \\ &\text{Continuum theory:} \qquad && \int d[\phi] e^{S(\mu)} = \lim_{\Lambda_0 \to \infty} \int d[\phi] e^{S_0(\Lambda_0) + \delta S(\mu,\Lambda_0)} \end{aligned} \tag{1.14}$$

with the same conventions as before ($\phi_{<\Lambda}$ refers to field configurations with energy lower than Λ). The effective RG transformation obtained by decreasing Λ is irreversible since it eliminates high-energy degrees of freedom, while the continuum RG transformation obtained by varying μ is reversible since it merely amounts to subtracting or adding some finite quantity in the action (i.e., to imposing a different renormalization condition).[24]

[23] For more technical details, see, e.g., Weinberg (1995, sec. 12.4; 1996, chap. 18), Zinn-Justin (2007), and Schwartz (2013, chap. 23). Here I rely on the standard understanding of the Wilsonian RG. For a formal interpretation, see Rosaler and Harlander (2019).

[24] Note the difference between active and passive transformation in both cases. An active RG transformation corresponds to a genuine change of scales (and hence to integrating out degrees of freedom in the effective case). A passive RG transformation corresponds to a conventional redefinition of energy units (in which case no degree of freedom is integrated out in the effective case).

Next, the most conceptually consistent interpretation of the cut-off and of the renormalization scale is not the same in the two cases. In the effective approach, the idea of integrating out *all* the high-energy degrees of freedom makes sense only if we use a sharp cut-off (e.g., a lattice or a momentum cut-off). If we use a smooth cut-off, the path integral measure is still defined by summing over arbitrarily high-energy states. Similarly, if we have good reasons to think that the high-energy states close to the sharp cut-off Λ_0 misrepresent, in some way or another, the correct state of matter, we should make sure that we exclude them. One conceptually simple and clear way of ignoring these high-energy states is to integrate out *all* the high-energy degrees of freedom between Λ and Λ_0. In contrast, the continuum approach is based on the idea that all the degrees of freedom in the original theory $\mathcal{L}$ are relevant in some respect. One way of making sure that the continuum assumption holds is to use a regularization method that gives weight to the physical states of interest without eliminating the others. In the method of dimensional regularization, for instance, the divergences are analyzed by shifting the dimension of space-time by $\pm\epsilon$, and the state space of the theory is smoothly distorted in the UV in a way that keeps all the possible energy states but significantly lowers the weight of the states above the scale μ.[25]

Agreed, one important lesson of the modern understanding of renormalization is that the specific value of the cut-off and the specific details of the regularization method do not really matter. They can always be absorbed in the formal expression of the renormalized parameters and, overall, the predictions obtained with different regularization methods are empirically equivalent. Yet, this does not mean that all regularization methods are on the same footing. If the goal is to define a theory across all energy scales, for instance, it appears somewhat conceptually inconsistent to construct the theory by first eliminating all the high-energy degrees of freedom beyond some fixed scale. Similarly, if the goal is to offer a restricted description of low-energy degrees of freedom, it appears somewhat conceptually inconsistent to include the contributions from arbitrarily

[25] In more detail, if we take $d = 4 \pm \epsilon$, we have to rescale each coupling by some power of the renormalization scale μ for dimensional consistency (e.g., replace λ by $\lambda\mu^{\epsilon}$) and modify the dimension of the divergent integrals. Integrating out the extra ϵ dimension in those integrals leaves an additional damping factor in the integrand that depends on both ϵ and μ. If we ignore potential trouble in the IR, this damping factor smoothly vanishes for momenta much larger than μ with ϵ small and μ fixed (Georgi, 1992, p. 4; 1993, sec. 2.4, provides a very helpful and concise explanation).

high-energy degrees of freedom when calculating low-energy predictions. At the very least, some regularization methods make these specific goals more explicit and provide a conceptually simpler and clearer way of achieving those goals. In the case of the effective approach, for instance, a deformation that eliminates all the high-energy degrees of freedom appears to be more natural than a deformation that merely lowers the weight of UV contributions. For if we believe that the theory literally breaks down at high energies, we should rather avoid using those high-energy degrees of freedom altogether instead misrepresenting their properties and using them to compute low energy predictions. Likewise, a sharp cut-off introduces a conceptually simple and clear classification of low-energy and high-energy field configurations, while a smooth cut-off makes the boundary between them somewhat fuzzier. Agreed again, both a sharp and a smooth cut-off offer a highly idealized representation of the boundary between the low-energy and high-energy regimes of the theory. But we do not need to take the exact form of the cut-off to be physically meaningful in order to grant that those differences between a sharp and a smooth cut-off regularization method are significant. And of course, if our primary goal is simply to compute low-energy predictions, we should probably select the regularization method which allows us to achieve this goal in the simplest, most efficient and appropriate way.

1.5.2 RG and Renormalizability

Now, let us look at the implications of the RG for the notion of renormalizability and for the scope of the continuum approach. Before we do that, it is necessary to spend some time clarifying: (i) the notion of RG space or theory-space; (ii) the notion of fixed point; and (iii) the different types of local behaviors in the vicinity of fixed points.[26]

(i) Consider first the infinite set of renormalized couplings $g_n(\Lambda)$, including both couplings from renormalizable and non-renormalizable interactions, which can be used to define any sort of renormalized (local) QFT in four dimensions for a specific set of fields and symmetries. The infinite set of RG equations span an infinite-dimensional space, the so-called “RG space", where

[26]The analysis does not depend on the type of RG used since the effective and the continuum RG are formally equivalent at the level of couplings.

each coupling stands for an independent coordinate and where each point in the space represents a theory defined at some energy scale Λ (see Fig. 1.3). The RG transformations of the couplings induce trajectories in this space, the so-called "RG flows", either towards the IR or the UV as we respectively decrease or increase the value of Λ.[27]

(ii) Fixed points g^* are defined by the points in theory-space where the RG flow terminates. The fixed point is either IR or UV depending on whether the RG flow converges to the fixed point in the low-energy or high-energy limit. In each case, the β-function $\beta(g_n^*)$ of each coupling vanishes at the fixed point, which means that each coupling $g_n(\Lambda) = g_n^*$ remains constant whether the value of Λ is increased or decreased and that the theory specified by the couplings g_n^* is scale-invariant. It turns out that we can distinguish between two kinds of fixed points. A Gaussian fixed point $g^* = 0$ defines a non-interacting theory, and theories converging towards a Gaussian fixed point are called "asymptotically free". A Wilson-Fisher (or non-Gaussian) fixed point $g^* \neq 0$ defines a non-trivial scale-invariant dynamics, and theories converging towards a Wilson-Fisher fixed point are called "asymptotically safe". As we can already anticipate, the existence of a UV fixed point indicates that the corresponding continuum theory behaves well at high energies, i.e., that the value of its couplings remains finite at high energies.[28]

(iii) The infinite set of RG equations determine local topological properties of the RG flow on theory-space. To see that, we need to examine first the behavior of couplings in the vicinity of a fixed point. In the simple case of a Gaussian fixed point, the perturbative RG equation for a coupling g looks like:

$$\Lambda \frac{\partial g}{\partial \Lambda} = \beta(g) \approx (-\Delta + \gamma)g + bg^2 + cg^3 + \mathrm{O}(g^4) \tag{1.15}$$

where γ, b and c are constants. Assuming that γ is negligible, the solution at lowest order is given

[27] Typically, an effective theory is defined by a finite segment of the RG flow with an upper bound while a (well-defined) continuum theory is defined by a complete segment specified by the values of $g_n(\mu)$ for any μ.

[28] More precisely, there are three conditions for asymptotic safety/freedom in both the IR and the UV case: (a) the vanishing of the β-function; (b) the existence of a finite-dimensional surface in the vicinity of the fixed point if we want the theory to be predictive; and (c) the existence of a well-behaved RG flow on the way to and at the fixed point. In the IR case, for instance, condition (c) obtains if $\sup_{\Lambda \leq \Lambda'} g(\Lambda) < \infty$ for some Λ' and $\lim_{\Lambda \to 0} g(\Lambda) = g* < \infty$.

by:

$$g(\Lambda) = (\frac{\Lambda}{\Lambda_0})^{-\Delta} g(\Lambda_0) \tag{1.16}$$

Three types of behaviors can be distinguished from this elementary perturbative equation, and each of them clarifies the scale-dependence of the non-renormalizable, renormalizable and super-renormalizable individual interaction terms defined in section 1.4 (here with the flow directed towards the IR):

(a) Super-renormalizable interaction: If $\Delta > 0$, the coupling $g(\Lambda)$ becomes large at small scales $\Lambda \ll \Lambda_0$ and negligible near the cut-off $\Lambda \lesssim \Lambda_0$. The coupling and the corresponding operator are said to be "relevant" at low energies.

(b) Non-renormalizable interaction: If $\Delta < 0$, the coupling $g(\Lambda)$ becomes negligible at small scales $\Lambda \ll \Lambda_0$ and large near the cut-off $\Lambda \lesssim \Lambda_0$. The coupling and the corresponding operator are said to be "irrelevant" at low energies.

(c) Renormalizable interaction: If $\Delta = 0$, dimensional analysis is ineffective. The sign of the next dominant term in Eq. 1.15 determines whether the coupling is "marginally" relevant or irrelevant. For instance, the (dimensionless) λ_R coupling in ϕ^4-theory is marginally irrelevant (see Eq. 1.13).

With these three properties in hand, we can specify the distinct local topological features of the RG flow in theory-space (see Fig. 1.3). The "critical surface" is defined by the set of couplings whose trajectory ends on the fixed point and the "unstable manifold" is defined by the set of couplings whose trajectory departs from the fixed point. In general, trajectories can be extremely varied: the flow might seemingly converge toward a fixed point and quickly diverge away as it comes too close to it, or the flow might seemingly diverge away from a fixed point and suddenly converge extremely fast towards it. Some RG flows even display periodic behaviors (see Wilson, 1971, and Bulycheva and Gorsky, 2014, for a discussion and examples of cyclic flows). In typical cases, the critical surface corresponds to the subspace spanned by the set of irrelevant couplings while the unstable surface corresponds to the subspace spanned by the set of relevant couplings, and most trajectories converge towards the fixed point and suddenly diverge away as the relevant couplings become too important (as depicted in Fig. 1.3). The analysis applies both to IR and UV fixed points, and we may speak similarly of IR/UV relevant, irrelevant and marginal operators.

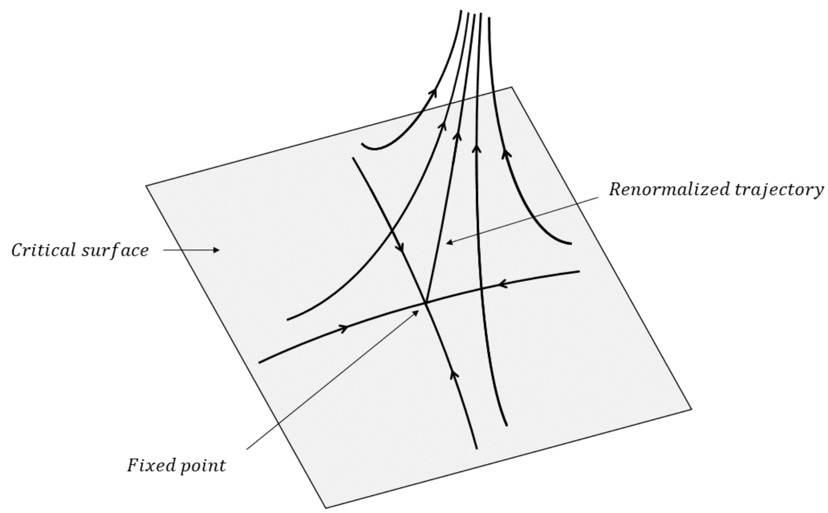

Figure 1.3: Theory-space in three dimensions, with a two-dimensional critical surface and a one-dimensional unstable manifold. The possible trajectories towards the IR are denoted by the lines with arrows.

This analysis has three important implications.[29] First, it shows that the pathological high-energy behavior of non-renormalizable interactions (i.e., IR-irrelevant/UV-relevant interactions) is closely linked to the fact that they generate increasingly divergent integrals in perturbation theory. Consider for instance a scalar theory in four dimensions with one non-renormalizable interaction term $g_6\phi^6$. The 6-particle physical amplitude $\Gamma(p_1, ..., p_6)$ is just equal to g_6 at first order. On dimensional grounds, the total amplitude Γ and any of the higher-order sub-amplitudes $g_6^n\Gamma_n$ ($n > 1$) have mass dimension -2. As I briefly indicated above, the amplitudes at some order n have in general the schematic form $g_6^n \int dk k^{D_n - 1}$. On dimensional grounds, we can therefore infer from the mass dimension of g_6^n (namely, $-2n$) that the number D_n increases with n ($D_n = 2n - 2$),

[29]Note that this analysis also explains why typical realistic QFTs are likely to be only approximately perturbatively renormalizable since they might contain IR irrelevant interactions that we have not detected yet (e.g., Butterfield and Bouatta, 2015; Williams, 2019a).

which shows that the sensitivity of non-renormalizable interactions to high energies is closely linked to the pathological perturbative behavior of the theory.

Second, the RG theory suggests a general non-perturbative characterization of renormalizability. In the continuum approach, the notion can be defined as follows (I drop the label "non-perturbative" for simplicity):

> A theory is renormalizable$_{RG}$ iff there is some μ' such that the RG flow remains on the same finite-dimensional surface as the theory is rescaled toward the UV (i.e., for any $\mu > \mu'$). (*mutatis mutandis* for non-renormalizable$_{RG}$.)

In other words, the theory is renormalizable$_{RG}$ if it can be expressed in terms of a finite number of independent couplings as the theory flows towards high energies and non-renormalizable$_{RG}$ if it is impossible to constrain the RG flow to stay on a finite subspace. If we add the requirement that the theory converges toward a UV fixed point g^* as μ is increased, we obtain Weinberg's characterization of renormalizability as asymptotic safety (Weinberg, 1979b, p. 802). The couplings of a theory satisfying this more sophisticated criterion of renormalizability, call it renormalizability$_{AS}$, remain finite at arbitrarily high energies. As we will see in section 1.6, this is a good sign that the predictions of the theory remain under good mathematical control at high energies.[30]

Third, the RG theory implies that the scope of the continuum approach is not as restricted as initially considered. The definition of relevant, irrelevant and marginal operators by means of dimensional analysis in Eq. 1.16 is only valid near a fixed point. In general, this is a good rule of thumb. But it is perfectly possible that some non-perturbative effects come into play either at low or high energies. In particular, it is perfectly possible that a coupling which looks UV-relevant at low energies actually happens to be well-behaved at high energies because of fortuitous non-perturbative quantum corrections. That is, a theory can both converge towards a UV Wilson-Fisher fixed point and include non-renormalizable interactions, as some physicists expect for the

[30] We can also speak of degrees of renormalizability$_{AS}$ or "approximate" renormalizability$_{AS}$ in the case of a renormalizable$_{AS}$ theory with additional UV-relevant interaction terms diverging at high energies if the contributions of these UV-relevant interactions are negligible compared to the contributions of the other interactions at low energies.

quantum-field-theoretic approach to quantum gravity.[31] Likewise, we might attempt to tame the pathological UV behavior of a given theory by embedding it into a larger theory displaying a UV fixed point. Overall, this suggests that the continuum approach is suitable for a larger class of physically relevant theories than initially expected. Yet, there still remains a large number of theories ill-handled under this approach, namely those which fail to converge towards a UV fixed point. As we will see in the next section, if we take the current perturbatively renormalizable formulation of the Standard Model by itself for instance, there are very good reasons to believe that it exhibits a Landau pole singularity and therefore makes inconsistent predictions at very high energies.

1.6 The Infinite Cut-Off Limit and the Continuum Limit

The goal of this section is to distinguish between different types of QFTs on the basis of their behavior in the continuum limit. For the sake of the argument, I will assume that the theory at hand has been renormalized under the continuum approach, except that we have kept the parameter Λ_0 fixed and not yet attempted to take the limit $\Lambda_0 \to \infty$. I will also assume that the analysis applies both to the perturbative and the non-perturbative case (with specific provisos when needed and with the speculative assumption that the non-perturbative notion of continuum QFT makes sense in realistic cases).

The first thing to note is that the notion of “continuum limit" is ambiguous. It may refer either to the *infinite cut-off limit* ($\Lambda_0 \to \infty$) or to the *continuum limit* ($\mu \to \infty$), properly speaking.[32] Note, moreover, that the distinction is robust under different regularization methods. For instance, the infinite cut-off and continuum limits correspond respectively to $\epsilon \to 0$ and $\mu \to +\infty$ for dimensional regularization. Likewise, using this terminology, taking the lattice spacing to zero in a lattice QFT amounts to taking the infinite cut-off limit, except in cases where the lattice spacing also plays the role of the renormalization scale (in which case there is only one type of limit).

[31] See Niedermaier and Reuter (2006) for a review of the asymptotic safety scenario.

[32] For a slightly different understanding of this distinction emphasizing the difference between the removal of a perturbative regulator and the removal of a non-perturbative regulator, see Delamotte (2012, sec. 2.6, esp. 2.6.3).

Accordingly, the notion of “good behavior" in the limit should be understood in two distinct ways:

(1) The low-energy predictions of the theory at μ are unaffected by taking the infinite cut-off limit $\Lambda_0 \to \infty$.

(2) The theory makes consistent predictions at arbitrarily high energies ($\mu \to \infty$).

(1) corresponds to cases where the low-energy physics described by the theory is sufficiently insensitive to the high-energy physics described by the theory, while (2) corresponds to cases where the high-energy predictions of the theory do not violate typical assumptions such as unitarity.[33] So (2) is not about empirical adequacy, properly speaking. After all, the theory might turn out to be empirically inaccurate at very high energies. But we should at least require that it makes consistent predictions, say, by making sure that the values of the couplings remain sufficiently small for unitarity to hold.

How should we discriminate between well-behaved and ill-behaved theories in the two cases then? Consider first the case of the infinite cut-off limit. Let us bracket any issue about the violation of the perturbative assumption in the case of the bare theory, and simplify the discussion by looking at the following toy-model theory $\mathcal{L}_R(\mu)$ with two renormalized couplings (g_0 and g_0' correspond to the bare couplings):

$$\begin{cases} g(\mu) = (\frac{\mu}{\Lambda_0})^{-\Delta} g_0(\Lambda_0) \\ g'(\mu) = (\frac{\mu}{\Lambda_0})^{\Delta} g_0'(\Lambda_0) \\ \Delta > 0 \end{cases} \tag{1.17}$$

The condition $\Delta > 0$ implies that g is super-renormalizable and g' non-renormalizable or, equivalently, that g and g' are respectively relevant and irrelevant near the IR Gaussian fixed point, and irrelevant and relevant near the UV Gaussian fixed point. In general, RG equations define families

[33] Unitarity is the assumption that the total sum of probabilities for the possible measurement outcomes of some specific physical process add up to unity.

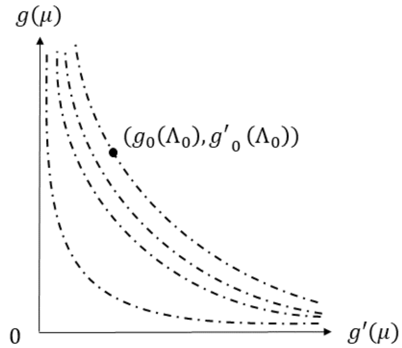

Figure 1.4: Theory-space in two dimensions with one IR-relevant/UV-irrelevant coupling g and one IR-irrelevant/UV-relevant coupling g'.

of solutions parametrized by different boundary conditions and, in the present case, each solution $(g(\mu), g'(\mu))$ of the two-dimensional RG equation is uniquely determined by specifying a single point $(g_0(\Lambda_0), g_0'(\Lambda_0))$ for some Λ_0 (see Fig. 1.4). Inversely, if the value of the renormalized couplings at μ is fixed by means of experiments, we can analyze the behavior of the bare theory in the infinite cut-off limit.

This toy-model is interesting because it displays two common types of behaviors. (i) $g'(\mu) = 0$ and $g(\mu) \neq 0$: that is, we do not detect non-renormalizable effects at low energies and take the liberty to fine-tune $g_0'(\Lambda_0)$ to zero, which implies $g'(\mu) = 0$. In this case, the RG flow lies on what is called the "renormalized trajectory" (see Fig. 1.3 and the $g(\mu)$ axis in Fig. 1.4) and we can take the infinite cut-off limit by assuming that the relevant bare coupling appropriately vanishes at infinity (i.e., such that $\lim_{\Lambda_0 \to \infty} \Lambda_0^\Delta g_0(\Lambda_0)$ is finite). It is therefore possible to take the infinite cut-off limit without affecting the low-energy predictions of the theory.

(ii) The most likely case is that both $g(\mu)$ and $g'(\mu)$ are non-zero, i.e., that the theory contains UV-relevant couplings. The toy-model indicates that the constraints we need to impose on these couplings are relatively minimal: the infinite cut-off limit leaves the low-energy predictions intact with the appropriate limits $g_0(\infty) = 0$ and $g_0'(\infty) = +\infty$. Of course, in general, taking the limit $\mathcal{L}_R(\mu) = \lim_{\Lambda_0 \to \infty}(\mathcal{L}_0 + \delta\mathcal{L})$ might require some delicate fine-tuning with the bare theory; and, as already emphasized, the perturbative assumption is explicitly violated in the case of the bare theory. But, overall, the RG theory is highly permissive since it is possible to take the infinite cut-off limit (at least formally) even if the theory contains pathological UV-relevant couplings. As we will see shortly, this fails to be the case if the renormalized coupling diverges at some finite energy scale Λ_M on the way to the limit.

Consider now the case of the continuum limit. RG flows towards the UV fall under four main types (e.g., Weinberg, 1996, sec. 18.3).[34] (i) Asymptotic freedom ($g^* = 0$) and (ii) asymptotic safety ($g^* \neq 0$) are the best case scenarios. In both cases, the values of the renormalized couplings remain finite in the continuum limit, which is a good sign that the theory makes consistent predictions at high energies since the main source of violations of (perturbative) unitarity comes from arbitrarily large values of the renormalized couplings in the expression of the scattering amplitudes. Of course, in those two cases as much as in the two cases below, our confidence in the behavior of couplings across energy scales depends on the reliability of the methods used to derive their expression. Asymptotic freedom is a special case in that respect. It is firmly based on perturbation theory like many of the results usually obtained from renormalization theory. But the fact that the values of the couplings become arbitrarily small at very high energies justifies the use of perturbation theory in the first place and suggests that non-perturbative results do not spoil the asymptotic behavior of the theory (e.g., a non-perturbative contribution to a scattering amplitude depending on some factor $e^{-1/g(\mu)}$ becomes arbitrarily negligible for $\mu \to \infty$ if $\lim_{\mu \to \infty} g(\mu) = 0$).

(iii) Let me call “asymptotic UV instability" the type of limiting behavior characteristic of theories containing divergent UV-relevant interactions as μ tends to ∞ (e.g., g' in the toy-model above). This case is problematic because, in general, these divergent UV-relevant interactions lead to violations of (perturbative) unitarity at high energies. These interactions even contain explicit information about the energy scale where those violations of unitarity arise. That being said, it is still possible to define the perturbative expression of the renormalized theory in the infinite cut-off limit $\Lambda_0 \to \infty$ if we restrict the range of the parameter μ to low energies. Moreover, if we use a smooth regularization method, the renormalized theory still includes negligible contributions

[34]Other cases include scale-invariance ($\beta(g) = 0$), in which case the RG does not flow, strictly speaking, and the cyclic behaviors mentioned above.

from arbitrarily high-energy excitation states (compared to contributions from low-energy states). Hence, even though the theory behaves badly in the continuum limit, the continuum assumption holds in this case for processes probed at sufficiently low energies.

(iv) Let me call “finite UV instability" the type of limiting behavior characteristic of theories containing a Landau pole, i.e., a finite energy scale Λ_M at which at least one of the couplings $g(\Lambda_M)$ diverges. As the ϕ^4-theory shows, finite UV instability is the worst case scenario. The solution to the perturbative RG equation of the quartic coupling $\lambda_R(\mu)$ is given by (cf. Eq. 1.13):

$$\lambda_R(\mu) = \frac{\lambda_R(\mu')}{1 - \frac{3\lambda_R(\mu')}{16\pi^2}\ln(\frac{\mu}{\mu'})} \tag{1.18}$$

Given a fixed experimental value $\lambda_R(E)$ at the energy scale E, the coupling λ_R diverges at $\Lambda_M = E\exp(16\pi^2/3\lambda_R(E))$. Similarly, if we evaluate the expression of the bare coupling at the scale $\mu = \Lambda_0$, the bare coupling diverges at the same finite scale Λ_M. And if we do not make any low-energy measurement and decide to take $\lambda_R(\mu) = 0$, we have to give up the initial assumption that the theory is an interacting theory. So, overall, the theory cannot be consistently defined in the infinite cut-off and continuum limits.[35] Now, the framework of the continuum approach is such that it is possible to take the infinite cut-off limit at the level of perturbative scattering amplitudes if we restrict ourselves to the first few orders in perturbation theory (see Eq. 1.8). However, the RG reveals that the partial perturbative relationship between the bare and the renormalized coupling obtained from the renormalization procedure is only superficially well-defined in the infinite cut-off limit: if we include the leading logarithms at higher orders in perturbation theory (as the derivation of the RG automatically does), we find a Landau pole.

In sum, continuum QFTs are likely to make consistent predictions at high energies when they are known with confidence to have a fixed point. The most reliable property of QFTs that we can typically find by means of perturbative methods is asymptotic freedom. And, for the large majority

[35] Of course, it might be the case that the Landau pole turns out to be an artifact of the perturbative analysis. For a discussion about the existence of a Landau pole and triviality in the case of QED, see, e.g., Gockeler et al. (1998a,b) and Gies and Jaeckel (2004).

of continuum QFTs, there are good reasons to believe that they are not only conceptually incoherent and physically dubious but also that they make inconsistent predictions at high energies—or, at the very least, that standard perturbative techniques cannot be used in those cases. Table 1.1 below summarizes the main interpretative differences between the effective and the continuum approach, including the results from section 1.6.

	Effective approach	**Continuum approach**
The continuum assumption	False	True
Goal	Select the appropriate low-energy degrees of freedom	Define the theory across all length scales
Bare theory	Physical theory	Intermediary/initial mathematical tool
Renormalized theory	Effective theory	Physical theory
Regulator Λ_0	The scale at which the theory breaks down	Intermediary mathematical tool
Regularization and renormalization method (most conceptually consistent)	Sharp cut-off (Λ)	Smooth cut-off (μ)
Infinite cut-off limit	Physically irrelevant	Mandatory
Continuum limit	Physically irrelevant	Consistent for a restricted class of well-behaved theories
Perturbative renormalizability	Perturbative predictions within ϵ with a finite number of parameters	Exact perturbative predictions with a finite number of parameters
Non-perturbative renormalizability	Finite-dimensional RG surface within ϵ + IR fixed point	Finite-dimensional RG surface + UV fixed point

Table 1.1: Main interpretative differences between the effective and the continuum approach.

1.7 Butterfield and Bouatta on Continuum QFTs

An advocate of the axiomatic approach might raise the following objection at this point: why should we take the differences between the effective and the continuum approach seriously if both fail to meet satisfying standards of mathematical rigor in the first place? And why should we attach any importance to the good behavior of asymptotically safe QFTs as opposed to finitely unstable QFTs if there is a chance that they are both mathematically inconsistent and *a fortiori* physically incoherent?[36] Wallace (2006, sec. 3-4; 2011, sec. 6-9) has rightly argued, I believe, that effective Lagrangian QFTs are as well-defined as any of the past theories that we usually take to be mathematically well-defined, and therefore should be considered fit for foundational and philosophical scrutiny. Butterfield and Bouatta (2014) recently extended this claim to continuum QFTs (see also Butterfield, 2014, pp. 8-9, sec. II.2-3, p. 31). They argue that even if the path integral formulation of realistic continuum QFTs has not yet received a precise mathematical definition according to the standards exhibited in the axiomatic, algebraic and constructive programs, some of these theories appear to be sufficiently mathematically well-defined according to physicists' common standards to be fit for philosophical scrutiny. Hence, by endorsing less stringent criteria of mathematical rigor, they claim, we should feel confident to draw world pictures for the heuristic formulation of *some* continuum QFTs. I will argue that the methodological and conceptual differences between the effective and the continuum approach discussed in sections 1.3-1.6 suggest reasons to temper Butterfield and Bouatta's claim.

Let me begin by making two friendly amendments to their discussion of continuum QFTs. (i) They contend that the contrast between theories likely to be (A) mathematically well-defined and (B) mathematically ill-defined depend, broadly speaking, "on the type of fields in the theory concerned" (Butterfield and Bouatta, 2014, p. 65). Agreed: as Butterfield and Bouatta rightly emphasize, in four dimensions, QFTs including only non-abelian gauge fields fall under case (A) while QFTs including only scalar or fermionic fields typically fall under case (B). In general,

[36]One might see these objections as particular cases of the general objection that the conventional mathematical apparatus of QFTs is ill-defined (e.g., Halvorson, 2007, p. 731; D. Fraser, 2008, p. 550; Kuhlmann, 2010; Baker, 2016, p. 5; Summers, 2016).

however, the field content of a QFT does not provide a reliable guide to assess whether the QFT is mathematically well-defined or not. Examples of asymptotically free scalar and fermionic QFTs in two and three dimensions show that the mathematical well-definedness of a QFT is not simply determined by the type of its quantum field operators.[37] In contrast, the scaling behaviors of QFTs exhibited by means of RG methods offer a more systematic way of distinguishing between (A) and (B), and Butterfield and Bouatta's diagnosis somewhat obscures the remarkable fact that this criterion does not depend on the content of the theory. Agreed, the definition of a particular RG space depends on the specification of a set of couplings and therefore on the specification of a set of (local) interactions—which, in turn, depends on the specification of a set of fields (e.g., scalar, fermionic, gauge, etc.), symmetries, and a space-time dimension. However, the possible types of RG trajectories, i.e., the possible types of behaviors of theories across energy scales, do not depend on these constraints. And so what it means for a theory to be mathematically well-defined is independent of the specific QFT model considered.

(ii) Butterfield and Bouatta's classification of QFTs under (A) and (B) is also incomplete. They argue that we should group asymptotically free, safe and conformal theories under (A) and theories presenting a Landau pole under (B). Agreed, this provides a good rule of thumb for the high-energy limit of continuum theories; and, for the perturbative theories we have so far, there are, in general, good reasons to expect that Landau poles in the IR ("infrared slavery") are perturbative artifacts, as it seems for perturbative QCD. However, it is worth being more systematic here since the non-perturbative definition of a theory might display, say, a Landau pole in the IR and asymptotic freedom in the UV. I distinguished in section 1.6 between (a) finite instability (i.e., existence of a Landau pole), (b) asymptotic instability (i.e., asymptotically divergent couplings), (c) asymptotic freedom (i.e., convergence to a zero fixed point), and (d) asymptotic safety (i.e., convergence to a non-zero fixed point), to which we might add the two additional cases of (e) non-convergent cyclic scaling behavior (i.e., non-convergent oscillating couplings) and (f) scale-invariant theories (i.e.,

[37] See, e.g., Weinberg (1996, sec. 18.3) and Gross (1999a, lecture 3, sec. 3.2, lecture 4). Examples of asymptotically safe theories in lower dimensions involving scalar or fermionic fields include: the Gross-Neveu model, the nonlinear σ-model with dimension $2 < d < 4$, and the 2d sine-Gordon model.

theories defined at a fixed point). It is perfectly possible that the non-perturbative definition of a theory displays two properties out of the five (a)-(e).

We should therefore only include under (A) theories defined by a continuous RG flow between two distinct fixed points and theories defined at a fixed point (ignoring (e)). The first class of theories corresponds to the class of IR/UV asymptotically safe/free theories, i.e., theories that continuously connect two conformal theories in the RG space.[38] For instance, the RG equation $\mu dg/d\mu = Ag - Bg^2$ for some coupling g with $A, B > 0$ describes the behavior of a theory asymptotically free in the IR (flowing towards the fixed point $g^* = 0$) and asymptotically safe in the UV (flowing towards the fixed point $g^* = A/B$). The second class of theories corresponds to the class of scale-invariant theories (i.e., $dg/d\mu = 0$). Although this has not been proven for models in dimension $d > 2$, these scale-invariant theories can typically be formulated as conformal field theories (CFTs).[39] Moreover, since our confidence in the existence of the properties (c), (d) and (f) is usually based on perturbative methods (as Butterfield and Bouatta rightly recognize), we should add the further constraint $g^* \ll 1$ for perturbation theory to be reliable.[40]

Let us now turn to Butterfield and Bouatta's claim that some continuum QFTs are ripe for metaphysical inquiry. At least as I understand them, Butterfield and Bouatta's claim relies on two key ideas. First, physics exhibits various standards of mathematical well-definedness and mathematical existence, and the heuristic standard commonly used in physics' practice provides a perfectly reasonable standard for interpretative purposes. In the context of QFT, the heuristic standard requires the theory to have a finite UV scaling behavior. By contrast, a theory is mathematically well-defined according to the axiomatic standard if it is axiomatizable and has a consistent model (Butterfield and Bouatta, 2014, p. 69). Second, the current perturbative formulation of some realistic continuum Lagrangian QFTs displays a UV fixed point and therefore satisfies the heuristic standard. QCD is one such example. Of course, the finite behavior of the theory at high energies

[38]For reference to the existence of well-defined and non-trivial RG flows from IR to UV fixed points, see, e.g., Caswell (1974), Banks and Zaks (1982), and Bond and Litim (2017).

[39]For references and discussions, see Polchinski (1988) and Dymarsky et al. (2015).

[40]Here it is worth mentioning the efforts made to formulate non-perturbative theories in the asymptotic safety scenario programme briefly mentioned in section 1.5.2.

does not mean that the functional integral resulting from the path integral quantization of the classical Lagrangian density is mathematically well-defined according to more stringent criteria of rigor. But the lack of a mathematically well-defined formulation should not prevent us from interpreting the heuristic formulation of our best continuum QFTs (Butterfield, 2014, p. 15). I take it that when Butterfield and Bouatta speak of the “heuristic" formulation of QFTs (Butterfield and Bouatta, 2014, p. 64, p. 68; Butterfield, 2014, p. 15), they refer to the current perturbative formulation that we have of these theories. And by ‘perturbative formulation’ I mean the formal expression of the path integral and the perturbative expression of the renormalized action and Lagrangian together with the set of perturbative techniques used to compute correlation functions.

Now, even if we accept to endorse less stringent criteria of mathematical rigor and philosophically assess the heuristic formulation of some continuum QFTs, it does not mean that we are warranted in attempting to draw “ontological claims" or “world-pictures" for these continuum QFTs (e.g., Butterfield and Bouatta, 2014, p. 68). It was central to the argument of section 1.3 that the structure of a physical theory does not only need to be under good mathematical control but also needs to make physical sense. Even if a QFT has a finite behavior at all energy scales, it is no indication that the theory has a physically coherent interpretation. Agreed, we do not need to demand that all the component parts of the theory make physical sense in order to dive into the metaphysical interpretation of a theory. But we should at least require that the core component parts of the theory do. Section 1.3 suggests that the perturbative formulation of our best continuum QFTs does not even meet this requirement in contrast to effective QFTs.

The argument goes as follows. To simplify the discussion and as already emphasized, I will follow Butterfield’s usage of the term ‘theory’ in its specific sense and identify the perturbative formulation of a QFT with the perturbative formulation of its Lagrangian (Butterfield, 2014, p. 31). Then, we may either interpret the renormalized Lagrangian or the bare Lagrangian (or both) in order to extract dynamical information. Consider first the renormalized Lagrangian. However we construct it, the renormalized Lagrangian together with the standard rules for deriving amplitudes yields divergent quantities if we do not restrict the state space of the theory. Hence, if the goal is to interpret empirically successful theories, we have no reason to even attempt to draw a world picture out of the renormalized theory or to take the renormalized Lagrangian to give us reliable dynamical information about the target system. At the very least, we should show some degree of caution.

Let us look at the bare Lagrangian. In the least naive perturbative construction of a renormalized continuum QFT, we start with some initial bare Lagrangian with the “wrong" parameters and we split it into a physical Lagrangian and a counter-term Lagrangian. The split is made in such a way that the counter-terms cancel the original divergences in the scattering amplitudes derived from the bare Lagrangian. And, by re-expressing the parameters of the bare Lagrangian, we find that the original bare amplitude is actually finite. The problem, however, is that the parameters of the bare Lagrangian diverge in the infinite cut-off limit. We precisely use the freedom that we have in defining the original bare parameters to absorb the UV divergences that we find in the original perturbative expansion. So, at least at this level, the original expression of the bare theory makes little physical sense. How about the “true" bare theory, i.e., the theory defined by the renormalized parameters evaluated at the cut-off Λ_0 (see section 1.3.2)? As already emphasized in section 1.3.3, there are concrete examples of theories where these bare parameters converge to a finite value in the infinite cut-off limit. However, if we choose to identify the bare parameters in this way, the resulting bare Lagrangian yields, once again, divergent predictions. Finite amplitudes are always derived from the theory which has the “wrong" parameters, as it were, since we always need to re-express the original couplings of the divergent amplitudes in order to absorb the divergences. And so whether we look at the renormalized or the bare Lagrangian, it does not appear that we can justifiably draw a world picture out of the perturbative formulation of a continuum QFT constructed under the continuum approach.

1.8 Conclusion

The aim of this chapter has been twofold: (i) to propose a general conceptual framework to understand the various aspects of renormalization theory based on the distinction between effective and continuum QFTs; and (ii) to show that the effective approach to renormalization offers a more physically perspicuous, conceptually coherent, and widely applicable way to construct perturbative QFTs in comparison to the continuum approach. The oddities of the continuum approach are best illustrated by the absence of physical justification for the introduction of counter-terms, the instrumental status of the bare theory, and the fact that, strictly speaking, the renormalized theory yields divergent amplitudes if we do not restrict the state space of the theory. Evaluating the limiting behavior of continuum QFTs also provides important conceptual and classificatory insights into the scope of the continuum approach: only asymptotically safe and free theories are likely to make consistent predictions at high energies in contrast to asymptotically and finitely unstable theories. In comparison, the effective approach is applicable to any local QFT model (as far as I am aware). The chapter concluded with some lessons for the debate over the interpretation of QFTs in response to Butterfield and Bouatta's paper (2014): the distinction between the effective and the continuum approach gives reasons to doubt that perturbative continuum QFTs are yet ripe for metaphysical analysis.

Chapter 2: In Praise of Effective Theories

Our best current theory of matter, the Standard Model (SM) of particle physics, is widely believed to be best formulated as an effective theory rather than as a putatively fundamental theory. The probable breakdown of the SM at short distances makes this view intuitively attractive insofar as an effective theory is designed to work well only within a limited domain. Yet, in light of the virtues and vices exhibited by past theories, effective theories also appear to be too *ad hoc* and complex to be even approximately true. My goal in this chapter is to clear the path for their epistemic appraisal. Using the example of the SM, I argue that these two vices are merely apparent.

2.1 Introduction

> The present educated view of the standard model, and of general relativity, is [...] that these are the leading terms in effective field theories. (Weinberg, 1999, p. 246)

There is a growing trend in the philosophy of physics to think that our most fundamental and empirically successful theories are best formulated as effective theories and thereby to align oneself with what has become, in Weinberg's words, the "educated" view in physics. This is intuitively attractive. The Standard Model (SM) of particle physics, for instance, is likely to break down at short distances and an effective theory is designed to work well only within a limited domain. The effective formulation of our best theories also exhibits many other virtues compared to their putatively fundamental formulation, including empirical fit and mathematical consistency (e.g., Cao and Schweber, 1993; Wells, 2012b), explanatory power (e.g., Hartmann, 2001), and robustness with respect to potentially new types of high-energy physics (e.g., Wallace, 2006, 2011; Williams, 2019b; J. D. Fraser, 2018, 2020b). And if these virtues are reliable indicators of approximate truth, as the historical record of physics suggests, we seem to be justified in thinking that our best effective theories offer a more reliable approximate picture of the world than their putatively fundamental counterparts and thus an epistemically privileged standpoint for engaging with foundational and interpretative matters.

There is a problem which, in my view, has not received enough attention: our best effective theories appear to be just too *ad hoc* and complex to be even approximately true. To be sure, some philosophers have not failed to point out the *ad hoc* character of the restriction—or “cut-off"—imposed on the domain of these theories, thereby suggesting that their success is better explained by some arbitrary tinkering than by their approximate truth (e.g., D. Fraser, 2009, 2011; Butterfield, 2014). In a similar vein, some philosophers speak of effective theories as “phenomenological" theories, thereby suggesting that these theories accommodate data in some regime without genuinely explaining them (e.g., Huggett and Weingard, 1995, p. 189; Butterfield, 2014, p. 65; see also Grinbaum, 2008, 2019). Yet, these criticisms are often too sketchy—and when they are not, little attention is paid to the details of realistic effective theories and to the variety and ambiguity of their vices. Effective theory supporters have, of course, responded to some of these criticisms (cf. Wallace, 2006, 2011; Williams, 2019b). But there are several issues which have not been fully addressed in my sense: for instance, whether the introduction of an arbitrary cut-off is still sufficiently well justified even if we concede that any appeal to new physics or to some putatively more fundamental theory is too open-ended to be significant (e.g., our best effective theories might not even be approximately derivable from the next theory).

The problem is all the more serious as there are good prospects for a distinct formulation of our best theories that does not exhibit those vices. The traditional continuum version of the SM is probably too mathematically ill-defined to present itself as a genuine alternative (cf., Wallace, 2006, 2011) and, even if we are comfortable with physicists’ standards of mathematical rigor, there is still clear evidence that this version cannot be consistently extended across all scales (see, e.g., Gies and Jaeckel, 2004). Yet the case does not seem to be entirely settled. Recent developments in the asymptotic safety program beyond the SM (and in quantum gravity) suggest that the SM might still be best formulated as a continuum theory applicable across all scales (and without the seemingly *ad hoc* restrictions and arbitrarily complex features displayed by its effective counterpart).[1]

Let me emphasize at the outset that, if this program were to be successfully completed, it would open up a serious case of transient underdetermination between two types of theories (or formulations). We can indeed express the various continuum and effective versions of the SM in the standard QFT language and adjust them without affecting their core theoretical content. So, strictly speaking, the underdetermination is neither between two different QFT frameworks nor between two particular QFT models in this scenario. Yet, taken literally, any continuum version of the SM and its effective counterpart make conflicting claims about the structure and content of the world. In particular, insofar as the effective version of interest makes inconsistent predictions at short distances, we cannot take it to approximately describe continuum fields or, for that matter, any kind of entities with a sufficiently fine-grained structure. We cannot even take it to literally describe in its own terms large distance patterns *of* such entities. By contrast, and insofar as the corresponding continuum theory makes consistent predictions across all scales, we can still take it to approximately describe continuum fields and large distance field patterns despite being incomplete (e.g., these continuum fields could still turn out to be "components" or "parts" of a more fundamental type of entity).

Now, current experimental data do not resolve by themselves whether the SM is ultimately best formulated as a continuum theory or as an effective theory. The conflict would also probably last longer than usual given the increasing difficulty of obtaining new data at high energies. And since we might have good reasons to prefer the effective formulation of the SM *even* if it can be consistently formulated as a continuum theory and embedded in some fundamental and complete theory, it does not seem that finding the successor of the SM would necessarily resolve the conflict. We

[1]See, e.g., Litim and Sannino (2014), Bond, Hiller, et al. (2017), Bond and Litim (2017), Bond, Litim, et al. (2018), Mann et al. (2017), and Eichhorn (2018, 2019). Of course, if these results are deemed too premature, the point would still be relevant, say, for the perturbative continuum formulation of Quantum Chromodynamics (QCD) and its Effective Field Theory (EFT) counterpart with higher-order non-renormalizable terms. Another option would be to look for candidates in the algebraic Quantum Field Theory (QFT) framework. But as rightly emphasized by Wallace (2006, 2011) in my view, the inability to formulate realistic algebraic models and derive from them predictions about, say, the branching ratio of the Higgs boson decay or the partial width of the Z boson decay in the case of the SM makes this option a non-starter, at least as of now. For a defense of the foundational and interpretative relevance of AQFT, see, e.g., D. Fraser (2009, 2011) and Kuhlmann (2010).

would still be left with the general question of whether non-fundamental and incomplete theories are best formulated as effective theories.[2]

My goal in this chapter is to defend the epistemic worth of effective theories through the example of the SM. I will not attempt to offer an exhaustive catalog of virtues and vices as a means to systematically compare effective and continuum theories and decide which one best accounts for current experimental data in the particular case of the SM (assuming that we do have a consistent continuum version of the SM). As I will briefly explain, the large amount of disagreement in the philosophy of science literature about the exact list of relevant virtues, their meaning, and their relative weight suggests that such a project is worthless. What is worthwhile, however, is to focus on specific vices which are taken to undermine the epistemic worth of effective theories. In what follows, I will argue that the vices of *ad hoc*ness and complexity are merely apparent. If we further assume that effective theories are overall more virtuous than their putatively fundamental counterparts, the argument clears the path for believing that the framework of effective theories is currently our most reliable technology for engaging with foundational and interpretative matters and that we should, in particular, take the SM to be best formulated as an effective theory.[3]

The chapter is organized as follows. Section 2.2 compares the continuum and effective formulations of the SM. For simplicity, I will use the traditional perturbatively renormalizable continuum version of the SM and the Standard Model Effective Field Theory (SMEFT), and pretend that the

[2]Note that the same issue would arise with General Relativity, although it is presumably harder to find a consistent continuum theory in this case since the Einstein-Hilbert action is already perturbatively non-renormalizable (see, e.g., Niedermaier and Reuter, 2006, for a comprehensive introduction to the asymptotic safety scenario in quantum gravity, and Eichhorn, 2019, for a recent report). Likewise, if we were to detect supersymmetric partners for the SM particles, the same issue would arise with the continuum and effective supersymmetric extensions of the SM (see, e.g., Bertolini, Thaler, and Thomas, 2013, for an EFT-friendly introduction to SUSY).

[3]Although I will not develop this point here, it is important to note that the framework of effective theories provides a new and relatively simple foundational scheme for understanding the “miracle" of physics, i.e., what makes its success possible in the first place, and which goes roughly like this: (i) use pragmatic, experimental, and historically motivated considerations to select an appropriate set of degrees of freedom and principles relevant in some regime; (ii) parametrize the effect of potentially new types of physics relevant in other unexplored regimes; (iii) evaluate the sensitivity of the physics within the regime of interest to the physics characterizing these other regimes; (iv) if you find a lack of sensitivity, use it to explain why the original choice of degrees of freedom and principles is epistemically and practically justified; (iv’) otherwise, rely on the pattern of sensitivity to justify the claim that there is something important missing in the original set of degrees of freedom and principles. (iv) is well illustrated by the continued use of Newtonian gravitation in the solar system and (iv’) by the naturalness problem in the SM as a key motivation for SUSY (see, e.g., S. Martin, 2016, pp. 3-5).

former remains consistent across all scales. Section 2.3 criticizes Wells's defense of the SMEFT. Section 2.4 examines the apparent *ad hoc*ness and complexity of this model. Section 2.5 concludes with some remarks about Gell-Mann's totalitarian principle, one of the key principles exemplified by effective theories.

2.2 Two Ways of Looking at the Standard Model

Despite enjoying a predictive accuracy defying common standards (up to ten parts in a billion in the electromagnetic sector) and having resisted several decades of attempts to find significant discrepancies in collider experiments, the SM is widely believed to be incomplete. On the experimental side, it does not account for the mass of neutrinos (inferred from atmospheric and solar fluxes), the different amount of protons and neutrons in the universe (inferred from the Cosmic Microwave Background), and the probable existence of "dark" matter (inferred from galactic rotation curves). On the theoretical side, most of the current contenders for a theory of quantum gravity imply that the SM (and more generally the QFT framework) becomes explanatorily and empirically deficient close to the Planck scale $l_p \sim 10^{-35}$ m (or, equivalently, close to the energy scale $E_p \sim 10^{16}$ TeV), if not much before. The SM also contains many features which, on the face of it, cry out for an explanation: why, for instance, does the SM contain nineteen independent parameters and why is its dynamics constrained by the particular gauge symmetry group $SU(3) \times SU(2) \times U(1)$? Whoever supposes that unification and simplicity are indispensable virtues of any complete theory of matter is likely to find the SM disappointing, to say the least.

Assuming, then, that there is a more fundamental, complete and empirically successful theory down the line, we might still wonder whether the SM is likely to remain approximately true and thus whether it is even worth interpreting in realist terms. As it turns out, the experimental anomalies mentioned above and current data obtained in collider experiments do not resolve by themselves whether the SM is best formulated as a renormalizable continuum QFT or as a non-renormalizable effective QFT. The particles of the SM are represented in both cases in terms of field-theoretic variables, say, a scalar field $H(x)$ defined at each point x of the Minkowski space-time for the Higgs boson. The dynamics of the SM is also encoded in both cases in a mathematical object called a Lagrangian density $\mathcal{L}$, which contains various types of operators describing the behavior of particles, say, a quartic self-interaction term $\lambda H^4(x)$ with coupling λ for the Higgs boson. Yet, the effective and continuum formulations of the SM still display incompatible features, and I will illustrate them by comparing the original version of the SM with the SMEFT.

2.2.1 The Standard Model

The traditional dynamics of the SM takes the following schematic form (see, e.g., Donoghue, Golowich, and Holstein, 1994; Burgess and Moore, 2006, for more details):

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{kinetic}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}} \tag{2.1}$$

where $\mathcal{L}_{\text{gauge}}$ and $\mathcal{L}_{\text{kinetic}}$ describe the behavior of the gauge bosons (i.e., photons, W and Z bosons, and gluons) and fermions (i.e., quarks and leptons) as well as their interactions, $\mathcal{L}_{\text{Higgs}}$ the behavior of the Higgs bosons and their interactions with the massive gauge bosons (i.e., W and Z bosons), and $\mathcal{L}_{\text{Yukawa}}$ the interactions between the Higgs bosons and fermions.[4] In total, $\mathcal{L}_{\text{SM}}$ contains twenty different kinds of Lagrangian operators (if we ignore hermitian conjugates) and nineteen free parameters (if we ignore neutrino masses and leptonic mixing angles).

Despite its apparent complexity, $\mathcal{L}_{\text{SM}}$ contains a relatively small number of operators constrained by a few core principles compared to its effective counterpart $\mathcal{L}_{\text{eff}}$ (cf. below). In particular, $\mathcal{L}_{\text{SM}}$ is invariant under Lorentz, gauge and CPT symmetry transformations (i.e., respectively: space-time translations, rotations and boosts; local phase transformations of the fermionic fields and gauge field transformations; and inversion of the charge sign and spatio-temporal orientations). $\mathcal{L}_{\text{SM}}$ also satisfies the key constraint of perturbative renormalizability. At a superficial level, this constraint simply means that $\mathcal{L}_{\text{SM}}$ includes operators of mass dimension four or less and therefore only a finite number of (linearly) independent operators since all the operators of $\mathcal{L}_{\text{SM}}$ must

[4] $\mathcal{L}_{\text{SM}}$ contains additional ghost and gauge-fixing terms once the theory is quantized via path integral methods.

have positive mass dimension.[5] Once these symmetry principles and constraints are enforced, the structure of the operators left is such that $\mathcal{L}_{\text{SM}}$ exhibits additional baryon and lepton number "accidental" symmetries (i.e., global transformations of the overall phase of the quark fields and the overall and individual phases of the lepton fields).

The constraint of perturbative renormalizability plays, in fact, a much more significant role than simply constraining the set of operators allowed in $\mathcal{L}_{\text{SM}}$. As is well-known, realistic QFTs in high energy physics suffer from a host of mathematical issues. For instance, if we attempt to directly confront $\mathcal{L}_{\text{SM}}$ to experiments, most of the quantities derived from the model by means of perturbative methods contain integrals that diverge in their high-energy domain of integration and give rise to inconsistent probabilistic predictions. The only solution found so far is to "renormalize" the model: for instance, by imposing an upper energy bound Λ on these integrals (i.e., a high-energy cut-off), absorbing their divergent Λ-dependent terms into the free parameters of the model, and canceling these terms by adjusting appropriately the value of the free parameters. Perturbatively renormalizable models have two remarkable features: (i) only a finite number of free parameters is required to absorb all the divergent terms (order by order in perturbation theory); (ii) once all the divergences have been absorbed, we can in principle take the cut-off Λ to infinity and consistently define the model with "renormalized" parameters across all energy scales, at least according to physicists' standards of mathematical rigor.

All of this would be appealing if the application of perturbative methods in QFT had not its own limitations. In particular, there are currently good reasons to believe that some of the renormalized parameters of $\mathcal{L}_{\text{SM}}$ still diverge at some finite high-energy scale and therefore that $\mathcal{L}_{\text{SM}}$ cannot be consistently defined across all scales (see, e.g., Gockeler et al., 1998a,b; Gies and Jaeckel, 2004). On the brighter side, recent work in the asymptotic safety scenario beyond the SM suggests that $\mathcal{L}_{\text{SM}}$ can be consistently defined across all scales with a minimal number of modifications and by keeping, in particular, only operators of mass dimension four or less.[6] The experimental anomalies

[5] The mass dimension Δ of a physical quantity is the power of that quantity expressed in terms of some energy variable (i.e., energy$^{\Delta}$) with natural units $c = \hbar = 1$.

[6] See, in particular, Litim and Sannino (2014) and Bond and Litim (2017) for generic results about asymptotic safety in gauge field theories with fermionic and scalar fields, and Bond, Hiller, et al. (2017) for a simple asymptotically safe

mentioned above can also be accommodated by including only these types of operators (see, e.g., Gouvea, 2016, for the issue of neutrino masses). And this suggests that the formulation of the SM as a (perturbatively) renormalizable continuum QFT is still very much a live option.

2.2.2 The Standard Model Effective Field Theory

We may also think that the probable explanatory and empirical failure of the SM at high energies makes it somewhat pointless to try to solve what appears to be a mere mathematical problem. If we keep the cut-off Λ fixed at some finite value, the original issue of divergences does not even arise, i.e., all the previous integrals are finite (ignoring potential divergences at low energies). Yet, once we compute these integrals, we find that the predictions of the model contain various types of Λ-dependent terms, and this might be deemed unsatisfactory since the cut-off is introduced "by hand" and its value somewhat arbitrary. To ensure that the predictions of the model do not depend on the cut-off, we need to absorb these terms into the free parameters of the model and cancel them just as before. However, the original set of operators is typically not sufficient to absorb the different types of Λ-dependent terms if we keep the cut-off fixed, i.e., we need to introduce new operators with free parameters. And if we wish to absorb *all* such terms instead, say, of ignoring the negligible ones in $1/\Lambda^n$ ($n \geq 1$) for sufficiently low energies compared to Λ, we typically need to introduce all the possible operators consistent with the symmetries of $\mathcal{L}_{\mathrm{SM}}$. As it turns out, we obtain the same result if we integrate out high-energy field configurations in the path integral formulation of the theory. The high-energy contributions to predictions above some finite cut-off typically generate all the possible dynamical terms structurally allowed by the original model at lower energies. And so if we have any good reason to impose a finite cut-off for a given model, this suggests that we should directly work with the model that includes all the operators allowed by its symmetries in the first place, i.e., before we even renormalize it (this point will be relevant in sections 2.4 and 2.5).

The Standard Model Effective Field Theory (SMEFT) is the simplest effective version of the

extension of the SM involving vectorlike fermionic fields.

SM obtained by taking this suggestion seriously, and its dynamics takes the following schematic form (see, e.g., Manohar, 2018; Brivio and Trott, 2019, for more details):

$$\begin{aligned}\mathcal{L}_{\text{eff}} &= \mathcal{L}_{\text{SM}} + \mathcal{L}^{(5)} + \mathcal{L}^{(6)} + ... \\ &= \mathcal{L}_{\text{SM}} + \frac{1}{\Lambda}\sum_k C_k^{(5)} Q_k^{(5)} + \frac{1}{\Lambda^2}\sum_k C_k^{(6)} Q_k^{(6)} + O(\frac{1}{\Lambda^3}) \end{aligned} \tag{2.2}$$

where the coefficients $C_k^{(i)}$ are dimensionless free parameters (the so-called "Wilsonian coefficients") and the (local) operators $Q_k^{(i)}$, which are organized here by increasing order of mass dimension $\Delta_i = i - 4$, involve various products of the fields of the SM and their derivative at the same space-time point. $\mathcal{L}_{\text{eff}}$ is no longer invariant under the accidental symmetries exhibited by $\mathcal{L}_{\text{SM}}$. Yet, $\mathcal{L}_{\text{eff}}$ still has the same particle content as $\mathcal{L}_{\text{SM}}$ and remains invariant under Lorentz, gauge and CPT transformations.

There are two crucial differences between these two simple continuum and effective models. First, $\mathcal{L}_{\text{eff}}$ contains an infinite sum of (linearly) independent operators suppressed by inverse powers of the cut-off Λ. At a superficial level, this means that $\mathcal{L}_{\text{eff}}$ fails to satisfy the constraint of perturbative renormalizability. At a deeper level, the higher-order operators $1/\Lambda^{i-4}\sum_k C_k^{(i)} Q_k^{(i)}$ typically generate violations of (perturbative) unitarity for energies of the order of the cut-off scale and thereby give rise to inconsistent predictions at high energies.[7] Hence, as a matter of principle, $\mathcal{L}_{\text{eff}}$ cannot provide reliable information about short distance features of the world and therefore about the structure of continuum fields compared to $\mathcal{L}_{\text{SM}}$. Second, $\mathcal{L}_{\text{eff}}$ contains an infinite number of independent parameters. Since we cannot in practice fix their value by collecting an infinite number of empirical inputs, there is no choice but to truncate $\mathcal{L}_{\text{eff}}$ at some finite order in $1/\Lambda$ if we are to make predictions at low energies $E \ll \Lambda$. The choice of a particular order depends on the number of available empirical inputs and some desired predictive accuracy. And if we were able to fix the infinite number of free parameters in experiments (or with the help of a more fundamental

[7] Unitarity is the assumption that the total sum of probabilities for the possible measurement outcomes of some specific physical process add up to unity. For a discussion about the intricate link between violations of perturbative unitarity and the onset of new physics, see, e.g., Aydemir, Anber, and Donoghue (2012) and Calmet and Casadio (2014).

theory), the most complete and exact version of the model would remain predictive and empirically accurate only up to some finite value of the scale Λ (apart from the exceptional case, for instance, where all the $C_k^{(i)}$'s are found to be zero).

Now, in principle, the exact form of the operators in $\mathcal{L}_{\text{eff}}$ is determined by a systematic algorithm constrained by the symmetries of the model. In practice, however, the task of specifying these operators is incredibly arduous. The order $i = 5$ is somewhat special. The symmetries of the model imply that there is only one operator in $O(1/\Lambda)$, the so-called "Weinberg operator", which involves products of the Higgs and lepton fields, breaks the lepton number symmetry, and generates neutrino masses (cf. Weinberg, 1979a). The situation becomes more complicated at order $i = 6$. The specification of a set of independent operators depends on a particular choice of operator basis. In the "Warsaw basis", which is commonly used in effective theories beyond the SM, the total number of operators in $O(1/\Lambda^2)$ which respect the original baryon number symmetry rises to 59 (if we ignore Hermitian conjugates, see Fig. 2.1 below), and up to 2499 if we distinguish between the quark and lepton fields (Abbott and Wise, 1980; Grzadkowski et al., 2010; Alonso et al., 2014). The determination of the number of operators (i.e., not their specific form) has only been achieved up to order $i = 8$ so far and it is expected that the number at each order grows exponentially (Lehman, 2014; Lehman and A. Martin, 2016; Henning et al., 2017). The point of all of this is of course that there is *no point* in trying to determine the exact form of the model. It is possible to summarize $\mathcal{L}_{\text{eff}}$ by means of an abstract analytic form involving an infinite hierarchy of operators. In practice, however, the increasing complexity of the higher-order operators makes the task of specifying the explicit form of the model impossible and actually useless if we are satisfied with the accuracy reached with the lowest-order terms.

2.2.3 And the winner is?

The SMEFT has at first sight little to commend it. The systematic expansion of $\mathcal{L}_{\text{eff}}$ in inverse powers of the cut-off scale appears to constitute an ingenious technique for organizing, predicting, and ultimately "saving" increasingly fine-grained phenomena—but not for getting even a glimpse at unobservable entities or structures.[8] Worse still, the exponentially growing complexity of $\mathcal{L}_{\text{eff}}$ suggests that we are pushing the unifying power of the QFT language to its limits when we attempt to formulate effective QFTs. In a word: $\mathcal{L}_{\text{SM}}$ easily wins.

An analogy might be helpful here. The increasingly large lists of independent operators at each order in the expression of $\mathcal{L}_{\text{eff}}$ have, as it turns out, something disturbingly akin to the mathematical tables that have been used by astronomers to register and further compute planetary positions up until the 20th century (see Figs. 2.1-2.2 below, and Norberg, 2003, for more details about these tables, often referred to as ephemerides). The mathematical language of the operators listed in Fig. 2.1 is obviously more advanced. These operators are also derived from more abstract principles than those used to derive the value of the orbital elements in Fig. 2.2 (namely, celestial Newtonian mechanics together with perturbative methods and the information contained in other ephemerides). Yet the two types of mathematical tools still have striking features in common. To mention only the three most important ones: (i) the table in Fig. 2.2 can be used to systematically classify the different types of phenomena associated with the orbit of a specific planet according to their relevance during a specific period of time in the same way the operators of $\mathcal{L}_{\text{eff}}$ can be used to systematically classify the different types of phenomena associated with some high-energy interaction process according to their relevance at some energy scale; (ii) once combined with a set of instructions or "precepts", the table can be used to compute the future positions of the planet up to some desired accuracy in the same way $\mathcal{L}_{\text{eff}}$ together with the standard perturbative methods can be used to compute probabilities for measurement outcomes up to some desired accuracy;[9] and (iii) both the size of the table and the number of operators in $\mathcal{L}_{\text{eff}}$ dramatically increase with the desired accuracy.

Now, the main role of the mathematical tables used in astronomy is to derive precise predictions in a systematic manner, not to account even partially for the complex causal structure underlying

[8] This instrumentalist interpretation actually fits EFT practitioners' own avowal that one of the main advantages of effective theories beyond the SM is that they allow us to accommodate and systematically organize the increasingly large and complex amount of data collected in current high-energy experiments (see, e.g., Brivio and Trott, 2019, pp. 40-41).

[9] Astronomers even used to call the combination of a set of tables and precepts a "theory" for the specific planet under investigation (Norberg, 2003, p. 180).

the orbits of planets. If we treat the SMEFT along the same lines, we might raise doubts as to whether the various higher-order operators in $\mathcal{L}_{\text{eff}}$ really stand for fine-grained features of new physics. These operators might simply be treated as mathematically efficient ways of systematically accommodating data beyond the SM. And if we follow this route, we might doubt that the negligible contributions of these operators to low-energy predictions reveal anything special about the robustness of the low-energy content of effective QFTs with respect to potentially new types of high-energy physics contrary to what Williams (2019b) and J. D. Fraser (2018, 2020b) claim for instance. In a word: if we take this analogy seriously, it seems that $\mathcal{L}_{\text{eff}}$ has no distinctive epistemic significance compared to $\mathcal{L}_{\text{SM}}$ and is primarily introduced for computational purposes.

2.3 Defending the Virtues of Effective QFTs *En Bloc*

As far as I am aware, Wells (2012a,b) is the only one who has attempted to systematically defend the epistemic worth of the SMEFT and it will be instructive to examine his strategy in some detail.

Wells first proposes to compare the respective virtues and vices of $\mathcal{L}_{\text{eff}}$ and $\mathcal{L}_{\text{SM}}$ along two different classifications. According to Ritcher's, which is supposed to illustrate a physicist's position on the matter, the epistemic worth of an empirically successful theory (or model) is best measured by its falsifiability and simplicity (Richter, 2006). Wells claims that $\mathcal{L}_{\text{SM}}$ wins on the count of falsifiability since $\mathcal{L}_{\text{eff}}$ contains an infinite number of free parameters which can be adjusted to accommodate new data. $\mathcal{L}_{\text{SM}}$ is also simpler since it contains only a finite number of independent terms and free parameters. As Wells (2012b, p. 65) rightly recognizes, the criterion of falsifiability is not always a reliable indicator of approximate truth—a radically false theory, for instance, might take many predictive risks. But he grants nonetheless that $\mathcal{L}_{\text{SM}}$ is overall more virtuous than $\mathcal{L}_{\text{eff}}$ on Ritcher's classification.

According to Thagard's (1978) classification, which is supposed to stand for a seminal philosophical view on the matter, the epistemic worth of an empirically successful theory is best measured by its ability to explain or unify diverse phenomena (consilience), the minimal number of

$1: X^3$	
Q_G	$f^{ABC} G_\mu^{A\nu} G_\nu^{B\rho} G_\rho^{C\mu}$
$Q_{\widetilde{G}}$	$f^{ABC} \widetilde{G}_\mu^{A\nu} G_\nu^{B\rho} G_\rho^{C\mu}$
Q_W	$\epsilon^{IJK} W_\mu^{I\nu} W_\nu^{J\rho} W_\rho^{K\mu}$
$Q_{\widetilde{W}}$	$\epsilon^{IJK} \widetilde{W}_\mu^{I\nu} W_\nu^{J\rho} W_\rho^{K\mu}$

$2: H^6$	
Q_H	$(H^\dagger H)^3$

$3: H^4 D^2$	
$Q_{H\Box}$	$(H^\dagger H)\Box(H^\dagger H)$
Q_{HD}	$\left(H^\dagger D_\mu H\right)^* \left(H^\dagger D_\mu H\right)$

$5: \psi^2 H^3 + \text{h.c.}$	
Q_{eH}	$(H^\dagger H)(\bar{l}_p e_r H)$
Q_{uH}	$(H^\dagger H)(\bar{q}_p u_r \widetilde{H})$
Q_{dH}	$(H^\dagger H)(\bar{q}_p d_r H)$

$4: X^2 H^2$	
Q_{HG}	$H^\dagger H\, G_{\mu\nu}^A G^{A\mu\nu}$
$Q_{H\widetilde{G}}$	$H^\dagger H\, \widetilde{G}_{\mu\nu}^A G^{A\mu\nu}$
Q_{HW}	$H^\dagger H\, W_{\mu\nu}^I W^{I\mu\nu}$
$Q_{H\widetilde{W}}$	$H^\dagger H\, \widetilde{W}_{\mu\nu}^I W^{I\mu\nu}$
Q_{HB}	$H^\dagger H\, B_{\mu\nu} B^{\mu\nu}$
$Q_{H\widetilde{B}}$	$H^\dagger H\, \widetilde{B}_{\mu\nu} B^{\mu\nu}$
Q_{HWB}	$H^\dagger \tau^I H\, W_{\mu\nu}^I B^{\mu\nu}$
$Q_{H\widetilde{W}B}$	$H^\dagger \tau^I H\, \widetilde{W}_{\mu\nu}^I B^{\mu\nu}$

$6: \psi^2 X H + \text{h.c.}$	
Q_{eW}	$(\bar{l}_p \sigma^{\mu\nu} e_r) \tau^I H W_{\mu\nu}^I$
Q_{eB}	$(\bar{l}_p \sigma^{\mu\nu} e_r) H B_{\mu\nu}$
Q_{uG}	$(\bar{q}_p \sigma^{\mu\nu} T^A u_r) \widetilde{H}\, G_{\mu\nu}^A$
Q_{uW}	$(\bar{q}_p \sigma^{\mu\nu} u_r) \tau^I \widetilde{H}\, W_{\mu\nu}^I$
Q_{uB}	$(\bar{q}_p \sigma^{\mu\nu} u_r) \widetilde{H}\, B_{\mu\nu}$
Q_{dG}	$(\bar{q}_p \sigma^{\mu\nu} T^A d_r) H\, G_{\mu\nu}^A$
Q_{dW}	$(\bar{q}_p \sigma^{\mu\nu} d_r) \tau^I H\, W_{\mu\nu}^I$
Q_{dB}	$(\bar{q}_p \sigma^{\mu\nu} d_r) H\, B_{\mu\nu}$

$7: \psi^2 H^2 D$	
$Q_{Hl}^{(1)}$	$(H^\dagger i \overleftrightarrow{D}_\mu H)(\bar{l}_p \gamma^\mu l_r)$
$Q_{Hl}^{(3)}$	$(H^\dagger i \overleftrightarrow{D}^I_\mu H)(\bar{l}_p \tau^I \gamma^\mu l_r)$
Q_{He}	$(H^\dagger i \overleftrightarrow{D}_\mu H)(\bar{e}_p \gamma^\mu e_r)$
$Q_{Hq}^{(1)}$	$(H^\dagger i \overleftrightarrow{D}_\mu H)(\bar{q}_p \gamma^\mu q_r)$
$Q_{Hq}^{(3)}$	$(H^\dagger i \overleftrightarrow{D}^I_\mu H)(\bar{q}_p \tau^I \gamma^\mu q_r)$
Q_{Hu}	$(H^\dagger i \overleftrightarrow{D}_\mu H)(\bar{u}_p \gamma^\mu u_r)$
Q_{Hd}	$(H^\dagger i \overleftrightarrow{D}_\mu H)(\bar{d}_p \gamma^\mu d_r)$
Q_{Hud} + h.c.	$i(\widetilde{H}^\dagger D_\mu H)(\bar{u}_p \gamma^\mu d_r)$

$8: (\bar{L}L)(\bar{L}L)$	
Q_{ll}	$(\bar{l}_p \gamma_\mu l_r)(\bar{l}_s \gamma^\mu l_t)$
$Q_{qq}^{(1)}$	$(\bar{q}_p \gamma_\mu q_r)(\bar{q}_s \gamma^\mu q_t)$
$Q_{qq}^{(3)}$	$(\bar{q}_p \gamma_\mu \tau^I q_r)(\bar{q}_s \gamma^\mu \tau^I q_t)$
$Q_{lq}^{(1)}$	$(\bar{l}_p \gamma_\mu l_r)(\bar{q}_s \gamma^\mu q_t)$
$Q_{lq}^{(3)}$	$(\bar{l}_p \gamma_\mu \tau^I l_r)(\bar{q}_s \gamma^\mu \tau^I q_t)$

$8: (\bar{R}R)(\bar{R}R)$	
Q_{ee}	$(\bar{e}_p \gamma_\mu e_r)(\bar{e}_s \gamma^\mu e_t)$
Q_{uu}	$(\bar{u}_p \gamma_\mu u_r)(\bar{u}_s \gamma^\mu u_t)$
Q_{dd}	$(\bar{d}_p \gamma_\mu d_r)(\bar{d}_s \gamma^\mu d_t)$
Q_{eu}	$(\bar{e}_p \gamma_\mu e_r)(\bar{u}_s \gamma^\mu u_t)$
Q_{ed}	$(\bar{e}_p \gamma_\mu e_r)(\bar{d}_s \gamma^\mu d_t)$
$Q_{ud}^{(1)}$	$(\bar{u}_p \gamma_\mu u_r)(\bar{d}_s \gamma^\mu d_t)$
$Q_{ud}^{(8)}$	$(\bar{u}_p \gamma_\mu T^A u_r)(\bar{d}_s \gamma^\mu T^A d_t)$

$8: (\bar{L}L)(\bar{R}R)$	
Q_{le}	$(\bar{l}_p \gamma_\mu l_r)(\bar{e}_s \gamma^\mu e_t)$
Q_{lu}	$(\bar{l}_p \gamma_\mu l_r)(\bar{u}_s \gamma^\mu u_t)$
Q_{ld}	$(\bar{l}_p \gamma_\mu l_r)(\bar{d}_s \gamma^\mu d_t)$
Q_{qe}	$(\bar{q}_p \gamma_\mu q_r)(\bar{e}_s \gamma^\mu e_t)$
$Q_{qu}^{(1)}$	$(\bar{q}_p \gamma_\mu q_r)(\bar{u}_s \gamma^\mu u_t)$
$Q_{qu}^{(8)}$	$(\bar{q}_p \gamma_\mu T^A q_r)(\bar{u}_s \gamma^\mu T^A u_t)$
$Q_{qd}^{(1)}$	$(\bar{q}_p \gamma_\mu q_r)(\bar{d}_s \gamma^\mu d_t)$
$Q_{qd}^{(8)}$	$(\bar{q}_p \gamma_\mu T^A q_r)(\bar{d}_s \gamma^\mu T^A d_t)$

$8: (\bar{L}R)(\bar{R}L) + \text{h.c.}$	
Q_{ledq}	$(\bar{l}_p^j e_r)(\bar{d}_s q_{tj})$

$8: (\bar{L}R)(\bar{L}R) + \text{h.c.}$	
$Q_{quqd}^{(1)}$	$(\bar{q}_p^j u_r) \epsilon_{jk} (\bar{q}_s^k d_t)$
$Q_{quqd}^{(8)}$	$(\bar{q}_p^j T^A u_r) \epsilon_{jk} (\bar{q}_s^k T^A d_t)$
$Q_{lequ}^{(1)}$	$(\bar{l}_p^j e_r) \epsilon_{jk} (\bar{q}_s^k u_t)$
$Q_{lequ}^{(3)}$	$(\bar{l}_p^j \sigma_{\mu\nu} e_r) \epsilon_{jk} (\bar{q}_s^k \sigma^{\mu\nu} u_t)$

Figure 2.1: Complete table of the six-dimensional operators Q_k preserving the baryon number symmetry in the SMEFT. The various letters X, L, R, ψ, G, W, B, H, l, e, q, u, d stand for various fields and products of fields, and the remaining letters for various indices, constants, and operators acting on other operators (from Alonso et al., 2014, p. 49; originally derived in Grzadkowski et al., 2010).

its auxiliary hypotheses, distinct kinds of entities and *ad hoc* assumptions (simplicity), and the amount of features it shares with other approximately true theories (analogy). Wells thinks that the competition between $\mathcal{L}_{\text{SM}}$ and $\mathcal{L}_{\text{eff}}$ is tighter in this case. The two models are equally consilient given current experimental data. $\mathcal{L}_{\text{SM}}$ wins again with respect to the criterion of simplicity. With respect to the criterion of analogy, however, Wells contends that $\mathcal{L}_{\text{eff}}$ wins since it fits better with the form of low-energy effective Lagrangians (e.g., Chiral Perturbation Theory) and the most com-

318 VENUS, 1855.

For Washington Mean Noon and Meridian Transit.

Mean Solar Time of Meridian Transit.	Sidereal Date.	Apparent Right Ascension. At Mean Noon.	Apparent Right Ascension. At Transit.	Apparent Declination. At Mean Noon.	Apparent Declination. At Transit.	Log. of Motion in a Sidereal Minute. In R.A.	Log. of Motion in a Sidereal Minute. In Dec.	Log. of Factor for Second Diff's. In R.A.	Log. of Factor for Second Diff's. In Dec.
d. h. m.		h. m. s.	m. s.	° ′ ″	′ ″				
Sept. 2 1 57.7	245	12 43 24.03	43 27.62	−11 5 32.1	6 47.9	+8.48371	−9.8077	−4.23	+5.06
3 1 54.4	246	12 44 4.34	44 7.27	11 20 32.6	21 42.2	8.40654	9.7838	4.25	5.08
4 1 51.1	247	12 44 37.41	44 39.68	11 34 42.8	35 46.4	8.30969	9.7576	4.26	5.10
5 1 47.6	248	12 45 3.04	45 4.67	11 48 0.1	48 57.7	8.18089	9.7280	4.27	5.12
6 1 43.9	249	12 45 21.04	45 22.05	12 0 21.9	1 13.5	7.99178	9.6948	4.27	5.14
7 1 40.2	250	12 45 31.24	45 31.67	12 11 45.4	12 31.0	+7.63752	9.6567	4.28	5.16
8 1 36.3	251	12 45 33.49	45 33.37	12 22 7.9	22 47.5	−7.08961	9.6132	4.28	5.18
9 1 32.2	252	12 45 27.67	45 27.03	12 31 26.6	32 0.3	7.83728	9.5624	4.29	5.20
10 1 28.1	253	12 45 13.66	45 12.54	12 39 38.7	40 6.6	8.10052	9.5023	4.29	5.22
11 1 23.8	254	12 44 51.36	44 49.80	12 46 41.4	47 3.9	8.26423	9.4293	4.29	5.24
12 1 19.3	255	12 44 20.74	44 18.79	12 52 32.0	52 49.3	8.38309	9.3380	4.30	5.26
13 1 14.7	256	12 43 41.77	43 39.51	12 57 7.9	57 20.2	8.47632	9.2177	4.30	5.27
14 1 10.1	257	12 42 54.50	42 51.99	13 0 26.5	0 34.2	8.55252	9.0430	4.30	5.29
15 1 5.3	258	12 41 59.00	41 56.30	13 2 25.1	2 28.7	8.61674	−8.7315	4.29	5.30
16 1 0.2	259	12 40 55.39	40 52.55	13 3 1.2	3 1.1	8.67159	+7.0457	4.28	5.31
17 0 55.1	260	12 39 43.86	39 40.96	13 2 12.6	2 9.2	8.71913	8.8035	4.27	5.32
18 0 49.9	261	12 38 24.63	38 21.75	12 59 57.5	59 51.3	8.76056	9.0947	4.25	5.33
19 0 44.6	262	12 36 58.00	36 55.21	12 56 14.3	56 6.0	8.79694	9.2691	4.23	5.33
20 0 39.0	263	12 35 24.31	35 21.67	12 51 2.2	50 52.5	8.82855	9.3940	4.20	5.34
21 0 33.4	264	12 33 44.07	33 41.66	12 44 20.6	44 10.2	8.85587	9.4912	4.16	5.33
22 0 27.8	265	12 31 57.80	31 55.69	12 36 9.7	35 59.5	8.87934	9.5701	4.12	5.33
23 0 22.1	266	12 30 6.09	30 4.34	12 26 30.7	26 21.2	8.89927	9.6358	4.06	5.32
24 0 16.2	267	12 28 9.60	28 8.26	12 15 25.1	15 17.2	8.91576	9.6920	3.99	5.31
25 0 10.2	268	12 26 9.06	26 8.19	12 2 55.5	2 49.9	8.92903	9.7396	3.90	5.29
26 0 4.3	269	12 24 5.21	24 4.84	11 49 4.9	49 2.3	8.93931	9.7810	3.79	5.27
26 23 58.3	270		21 58.99		33 58.3				
27 23 52.3	271	12 21 58.84	19 51.48	11 33 57.2	17 42.5	8.94656	9.8170	3.61	5.24
28 23 46.2	272	12 19 50.79	17 43.18	11 17 37.0	0 19.8	8.95074	9.8480	−2.59	5.21
29 23 40.2	273	12 17 41.94	15 34.96	10 60 9.4	41 55.7	8.95193	9.8748	+2.58	5.17
30 23 34.1	274	12 15 33.19	13 27.75	10 41 40.0	22 36.7	8.95004	9.8978	3.59	5.13
Oct. 1 23 28.1	275	12 13 25.46	11 22.41	10 22 15.3	2 29.7	8.94497	9.9172	3.67	5.07
2 23 22.2	276	12 11 19.65	9 19.82	9 62 2.3	41 42.3	8.93677	9.9332	3.83	4.99
3 23 16.2	277	12 9 16.62	7 20.70	9 41 8.8	20 22.3	8.92562	9.9460	3.93	4.89
4 23 10.4	278	12 7 17.12	5 25.78	8 79 42.7	58 37.2	8.91143	9.9555	4.01	4.78
5 23 4.7	279	12 5 21.89	3 35.79	8 57 51.7	36 34.8	8.89395	9.9622	4.08	4.61
6 22 59.1	280	12 3 31.66	1 51.46	8 35 43.5	14 22.5	8.87281	9.9665	4.14	4.36
7 22 53.5	281	12 1 47.15	0 18.41	7 73 25.7	52 7.8	8.84751	9.9685	4.19	+3.68
8 22 48.1	282	11 60 9.05	58 42.21	7 51 5.9	29 58.1	8.81794	9.9681	4.23	−4.12
9 22 42.8	283	11 58 37.84	57 18.30	7 28 51.6	8 0.6	8.78390	9.9652	4.25	4.49
10 22 37.6	284	11 57 14.01	56 2.07	6 66 50.1	46 22.2	8.74462	9.9598	4.27	4.68
11 22 32.5	285	11 55 57.93	54 53.87	6 45 8.3	25 9.0	8.69933	9.9521	4.29	4.80
12 22 27.6	286	11 54 49.97	53 54.01	6 23 52.3	4 26.8	8.64639	9.9424	4.30	4.88
13 22 22.9	287	11 53 50.45	53 2.74	5 63 8.0	44 20.7	8.58389	9.9305	4.31	4.95
14 22 18.3	288	11 52 59.58	52 20.22	5 43 0.5	24 55.5	8.50872	9.9163	4.32	5.00
15 22 13.8	289	11 52 17.54	51 46.58	5 23 3[illegible].6	6 15.7	8.41660	9.8999	4.33	5.05
16 22 9.5	290	11 51 44.44	51 21.86	4 64 54.7	48 25.0	8.29816	9.8812	4.34	5.08
17 22 5.3	291	11 51 20.33	51 6.10	4 47 4.6	31 26.5	8.13412	9.8603	4.34	5.11
18 22 1.2	292	11 51 5.23	50 59.28	4 30 7.3	15 23.1	7.86736	9.8370	4.34	5.13
19 21 57.4	293	11 50 59.12	51 1.38	4 14 5.8	0 17.4	−7.05383	9.8113	4.33	5.15
20 21 53.6	294	11 51 1.98	51 12.35	3 59 2.6	46 11.2	+7.70615	9.7828	4.33	5.16
21 21 50.0	295	11 51 13.74	51 32.06	3 44 59.5	33 6.0	8.05008	9.7514	4.33	5.17
22 21 46.6	296	11 51 34.28	52 0.35	3 31 57.9	21 3.1	8.23714	9.7169	4.32	5.18
23 21 43.2	297	11 52 3.43	52 37.07	3 19 58.9	10 3.3	8.36526	9.6791	4.31	5.18
24 21 40.0	298	11 52 41.02	53 22.04	3 9 3.6	0 6.9	8.46209	9.6371	4.30	5.18
25 21 37.0	299	11 53 26.86	54 15.07	2 59 12.1	51 14.3	8.53983	9.5807	4.29	5.19
26 21 34.1	300	11 54 20.79	55 15.95	2 50 24.6	43 25.4	8.60417	9.5365	4.28	5.19
27 21 31.3	301	11 55 22.58	56 24.44	2 42 41.3	36 39.7	8.65854	9.4766	4.27	5.19
28 21 28.6	302	11 56 31.97	57 40.32	2 36 1.5	30 56.7	8.70544	9.4079	4.26	5.18
29 21 26.0	303	11 57 48.74	59 3.36	2 30 24.4	26 15.7	8.74646	9.3274	4.25	5.18
30 21 23.7	304	11 59 12.68	60 33.34	2 25 49.4	22 35.8	8.78305	9.2284	4.23	5.17
31 21 21.4	305	12 0 43.56	2 10.03	− 2 22 15.6	19 56.1	+8.81583	+9.1034	+4.21	−5.16

Figure 2.2: The Ephemeris of Venus for September and October 1855 (reproduced in Campbell-Kelly et al., 2003, p. 176; originally from the *American Ephemeris and Nautical Almanac*, United States Naval Observatory. Nautical Almanac Office, 1855, p. 330).

plete forms of the theories which have been superseded so far (e.g., the Newtonian gravitational theory with relativistic correction terms derived from classical General Relativity). Thagard's classification, Wells concludes, does not favor either model. But if we "average" over Ritcher's and Thagard's classifications, $\mathcal{L}_{\text{SM}}$ appears to be slightly better than $\mathcal{L}_{\text{eff}}$.

The core of Wells's argument is that this type of epistemic assessment is misleading insofar as it fails to take into account the relative importance of the various criteria used to evaluate the two models. Wells grants that $\mathcal{L}_{\text{eff}}$ is less simple and falsifiable. But he also claims that these defects are counterbalanced by other non-negotiable virtues: namely, "observational consistency" and "mathematical consistency" (Wells, 2012b, p. 69).[10] By 'observational consistency' he means that the observational consequences of a theory are "consistent with all known observational facts" (2012b, p. 68). Wells does not explain why he thinks that $\mathcal{L}_{\text{eff}}$ wins on this count. But he presumably takes $\mathcal{L}_{\text{eff}}$ to have more extensive and appropriate resources to account for experimental discrepancies. On the other hand, Wells does not explicitly define the notion of mathematical consistency but uses instead several examples to illustrate it, including the fact that an explicit mass term in the Lagrangian density of a pure gauge theory is "mathematically inconsistent" with the invariance of the Lagrangian under gauge transformations (2012b, p. 68). And here again, Wells takes $\mathcal{L}_{\text{eff}}$ to win:

> The claim behind the ascendancy of effective theories is that unless there is good and explicit reason otherwise, consistency requires that a theory have all possible interactions consistent with its symmetries at every order. (2012b, p. 69)

Overall, then, $\mathcal{L}_{\text{eff}}$ prevails over $\mathcal{L}_{\text{SM}}$ because it is better equipped to accommodate current and future data and because it includes all the possible operators allowed by the core principles of the original model.

Wells's account still suffers from severe issues. Suppose first that we grant that either Ritcher's

[10]For a similar strategy in the philosophy of science literature, see, e.g., Douglas's (2009; 2013) distinction between "minimal criteria", i.e., necessary criteria for taking a theory to be approximately true, and "ideal desiderata", i.e., criteria which reinforce our confidence in the approximate truth of a theory but which might be found wanting without affecting its epistemic worth.

or Thagard's classification picks an exhaustive list of decisive epistemic virtues for theory-assessment and that the meaning of these virtues is unambiguous. Then, the case for $\mathcal{L}_{\text{eff}}$ is actually much worse than Wells thinks. Since there are good reasons to believe that both $\mathcal{L}_{\text{SM}}$ and $\mathcal{L}_{\text{eff}}$ are incomplete, the crucial question is whether they are falsifiable at sufficiently low energies (compared, say, to the Planck scale). And the issue with $\mathcal{L}_{\text{eff}}$ is not merely that it is less falsifiable than $\mathcal{L}_{\text{SM}}$ at low energies but also that it becomes predictively powerless if we try to make it as "observationally consistent" as possible. Strictly speaking, the most complete version of $\mathcal{L}_{\text{eff}}$ does not even make *any* prediction since it is impossible to fix the value of an infinite number of free parameters with experimental inputs. Concerning the criterion of analogy, Wells begs the question. He assumes that $\mathcal{L}_{\text{eff}}$ shares more features with other approximately true models than $\mathcal{L}_{\text{SM}}$. But this requires having justified that incomplete or non-fundamental theories are best formulated as effective theories in the first place.

Suppose now that we acknowledge the importance of distinguishing between negotiable and non-negotiable criteria. One might still wonder whether Wells's non-negotiable criteria are sufficiently clear and distinct to do the job he wants them to do. Take observational consistency for instance. Wells rightly concedes that we do not need to have tested all the possible observational consequences of a theory or that a theory does not need to fit all available data in order to show that it is more observationally consistent than its competitors. But even if we grant this point, the notion of "observational consistency" is still too vague to support $\mathcal{L}_{\text{eff}}$ against $\mathcal{L}_{\text{SM}}$. For instance, we might distinguish between predictive accuracy (numerical agreement between the predictions of the theory and available data), predictive competency (the ability of the theory to fit with available data), predictive quality (its ability to fit with a large variety of relevant data), predictive resilience (its ability to fit with new data), predictive novelty (its ability to make new predictions compared to previous theories), and predictive power (the ratio of independent predictions over the number of free parameters). Some of these virtues are certainly negotiable if the theory displays other kinds of virtues, such as explanatory power and simplicity. For instance, since the SM is likely to be superseded one day, we should probably take the predictive accuracy of a model in this context to be less significant than its ability to explain phenomena within its domain in simple terms. We should also probably take the decreasing predictive power of $\mathcal{L}_{\text{eff}}$ to be more significant than its increasing predictive accuracy as we include increasingly many operators. And, of course, the same sorts of issues affect Wells's notion of "mathematical consistency".[11]

Suppose at last that we grant that the criteria of observational consistency and mathematical consistency are sufficiently clear and distinct. Wells's claim that these virtues counterbalance the defects of effective QFTs still remains too elusive in the absence of an exhaustive list of virtues and vices with specific weights. Otherwise, we might wonder whether he has forgotten another non-negotiable virtue or vice which tips the scale in favor of $\mathcal{L}_{\text{SM}}$. Discussions in the philosophy of science literature about the importance of theoretical virtues (or super-empirical virtues, or complementary virtues) in matters of epistemic assessment (or theory-choice, or theory-appraisal, or theory-acceptance) have a long and winding history. What is striking is the amount of disagreement about what constitutes an exhaustive list of relevant virtues for assessing the epistemic worth of theories, even in well-delineated contexts (see, e.g., Kuhn, 1977; Newton-Smith, 1981, chap. IX, sec. 8; McMullin, 1996; Keas, 2018; Schindler, 2018). In addition to empirical fit, consistency and coherence, philosophers and historians of science often appeal to fertility, fruitfulness, scope, generality, explanatory power, unifying power, naturalness, durability, robustness, structural continuity, simplicity, parsimony, and elegance (among other virtues).

I am not going to attempt to define all these virtues. Suffice it to say that the list can probably be extended *ad nauseum* by distinguishing between more fine-grained virtues (as in the case of observational and mathematical consistency) and including particular virtues made vivid in specific contexts (see Wells, 2018, for an attempt to draw a relatively exhaustive list in the context of high energy physics). The list can also probably be reshuffled in various ways depending on how we understand controversial virtues such as "explanatory power" and "simplicity" and on how we classify redundant ones (e.g., the ability to make novel predictions as an empirical virtue or as a

[11] In particular, it does not account for the difference between logical consistency, external coherence, mathematical well-definedness, and lack of *ad hoc* assumptions (among other virtues), and even the criteria of logical consistency and mathematical well-definedness appear to be negotiable in actual physics practice (see, e.g., Meheus, 2002; Davey, 2003; Vickers, 2013, for a discussion).

specific type of fertility). Assigning specific weights to each virtue and deciding whether some virtue is truly epistemic or merely pragmatic is likely to be controversial too, not to mention that some virtues appear to be sometimes in conflict with one another (e.g., empirical fit and simplicity). So not only does it appear to be hard to maintain that observational and mathematical consistency is non-negotiable, or even intrinsically truth-conducive for that matter (see Wells, 2012b, p. 68, footnote 1, for his insistence on the epistemic notion of "best"). It is also probably worthless to try to give an exhaustive list of relevant epistemic virtues and assign specific weights to them.

Wells is right to emphasize that there is no choice but to rely on the properties of competing theories (or models) if we are to decide between them, whether we take the historical record of successful theories to make some properties more truth-conducive than others (see, e.g., Newton-Smith, 1981, pp. 225-226; Schindler, 2018, chap. 1) or use these properties in meta-inductive arguments to isolate the appropriate theory (see, e.g., Dawid, 2013; Castellani, 2019, p. 176). The stakes are high too, especially when it comes to allocating resources for constructing experimental devices which are more likely to provide the relevant test for one specific theory as opposed to others. For instance, we might think that building increasingly large particle colliders to test small experimental deviations captured by $\mathcal{L}_{\text{eff}}$ is wrong-headed and that we should rather look for signatures of quantum gravity in gravitational wave experiments with the hope of restricting the set of models in which $\mathcal{L}_{\text{SM}}$ could be exactly embedded. Be that as it may, Wells's insistence that the observational and mathematical consistency of effective QFTs counterbalances their vices appears to be misguided, and his overall strategy unlikely to succeed.[12]

To be fair, effective QFTs do exhibit relatively unambiguous virtues (see, e.g., Cao and Schweber, 1993, sec. 3; Hartmann, 2001). For instance, and as already emphasized, the structure of effective QFTs makes them easily fit with new empirical data and maximize their predictive accuracy in limited regimes. The introduction of a cut-off easily solves some of the most pressing mathematical issues underlying realistic QFTs. Since $\mathcal{L}_{\text{eff}}$ has the same particle content and core

[12] A similar conclusion applies to the various attempts made in the literature to elevate consistency and mathematical well-definedness as central criteria for selecting theories worthy of foundational scrutiny (see, e.g., D. Fraser, 2009; Kuhlmann, 2010).

symmetries as $\mathcal{L}_{\text{SM}}$, we might also expect that $\mathcal{L}_{\text{eff}}$ has appropriate explanatory resources to account for low-energy phenomena. The higher-order operators of $\mathcal{L}_{\text{eff}}$ also make it extremely fruitful for deriving constraints on potentially new high-energy phenomena. And, to give one last and perhaps more controversial example, effective QFTs even appear to be beautiful in the eyes of some EFT practitioners (cf. Shankar, 1999, pp. 54-55).

The real difficulty is whether the apparent *ad hoc*ness and complexity of effective QFTs do not undermine the overall value of these virtues. The vice of *ad hoc*ness is often taken to signal that the arbitrary intervention of some theorist explains better the success of a model (or hypothesis) than its approximate truth. Likewise, the vice of complexity is often taken to signal that a model (or hypothesis) is introduced to accommodate a complex data set rather than to genuinely explain it. The next section assesses whether effective QFTs actually exhibit these two vices (instead of attempting to make any sort of systematic epistemic balance sheet as explained above).

2.4 Too *Ad Hoc* and Complex to be True?

Effective QFTs have been mainly criticized for their *ad hoc*ness and complexity (e.g., Redhead, 1999; Buchholz, 2000; D. Fraser, 2006, 2008, 2011, 2018; Butterfield, 2014; Butterfield and Bouatta, 2015). Yet, the criticisms raised are often too sketchy to make it clear whether there is anything intrinsically wrong with these theories. Consider Redhead as a warm-up example. He briefly mentions that the EFT program requires us to give up "the search for an ultimate underlying order characterized by simplicity and symmetry" and "retreat to a position that is [...] somehow less intellectually exciting" (Redhead, 1999, p. 40). But if the point is merely that a never-ending succession of increasingly comprehensive effective theories with infinitely many terms at each stage is aesthetically unpleasing, we might wonder whether we are not treating these theories unfairly. On the other hand, philosophers and physicists who defend effective QFTs often say too little about their *ad hoc*ness and complexity to make it clear whether there is not some way in which their epistemic worth is undermined by these two vices (e.g., Hartmann, 2001; Wallace, 2006, 2011; Wells, 2012b; Williams, 2019b). I will now argue that the *ad hoc*ness and complexity of effective QFTs are merely apparent.

2.4.1 *Ad hoc*ness

Effective QFTs display at first sight a variety of *ad hoc* features. The most discussed in the literature is the introduction of a finite cut-off in the formulation of our best current effective QFTs (e.g., D. Fraser, 2006, p. 160; 2008, pp. 552-553; Strocchi, 2013, p. v; Butterfield, 2014, p. 40; Butterfield and Bouatta, 2015, pp. 23, 38). For instance, D. Fraser writes:

> [...] a compelling argument against relying on the cutoff representation is that introducing cutoffs is an ad hoc solution to the problem of infinite renormalization. We are not justified in introducing the assumptions that space is discrete and all systems are finite in spatial extent solely because it is a simpler means of achieving the end goal of obtaining a mathematically well-defined and consistent representation. (2006, p. 160)

Here D. Fraser assumes that effective QFTs need to be defined on a lattice of finite spacing and extent and include only a finite number of degrees of freedom in order to fit the standards of mathematical well-definedness exhibited in the axiomatic, algebraic and constructive programs (see, e.g., Summers, 2016). In a similar vein, although somewhat more cautiously, Butterfield writes:

> Here I will develop one position, often called the *effective field theory program* (or: approach). It is based not on confidence about the two topics above [viz. using QFT at high energies and accepting results obtained from a heuristic formalism rather than by rigorous mathematical proofs] but on an opportunistic or instrumentalist attitude to being *unconfident* about them" (2014, p. 40; see also Butterfield and Bouatta, 2015, p. 23)

So the worry, made especially clear in D. Fraser's quote, is that the finite cut-off Λ in $\mathcal{L}_{\text{eff}}$ is introduced *just* to solve the embarrassing mathematical issues exhibited by $\mathcal{L}_{\text{SM}}$ at high energies. We may think that the probable breakdown of the entire QFT framework at high energies and the experimental anomalies mentioned in section 2.2 provide good reasons to impose explicit restrictions on the domain of the SM. We are simply wrong: these sorts of reasons, or so the argument goes, are too open-ended to justify the introduction of an arbitrary finite cut-off.[13]

Of course, if the introduction of a finite cut-off or “cut-off hypothesis" in short happens to be justified by some other means, we can treat the simplicity and efficiency of this solution as a side-benefit rather than as its main rationale. Suppose, then, that we grant that the cut-off hypothesis needs to be assessed independently of speculative considerations arising from quantum gravity and experimental anomalies (cf., e.g., D. Fraser, 2009, p. 561). There are still at least two comparatively good reasons to keep a finite cut-off instead of taking it to infinity.

The first, which resonates with Butterfield’s quote above, is a matter of epistemic modesty. Why should we believe that our best models at a given time will remain reliable across all scales in the first place? There does not seem to be any good reason to be rather confident than unconfident. The predictive success of a model in a limited regime supports the claim that the model reliably accounts for phenomena in that regime, not that it will remain reliable across all regimes (or, for that matter, become unreliable in some unexplored regime). Without further empirical evidence, we are rather justified in restricting the scope of the model to that regime instead of extending it to all regimes. And, at least at first sight, appealing to the foundational or interpretative virtues of putatively fundamental models should not affect this conclusion insofar as we aspire to use our most reliable models for engaging with foundational and interpretative matters in the first place.

The second is a matter of instrumental conservatism. The repeated predictive failure of our best past models at some scale provides us with good reasons to believe that the current ones will endure the same fate (independently of whether they are approximately true and of whether there is ultimately some maximally empirically adequate model). And since there is nothing intrinsic to $\mathcal{L}_{\text{SM}}$ which suggests that it has some special status compared to its ancestors, we seem to have

[13] D. Fraser (2008, pp. 552-553) also claims that it is “illegitimate" to appeal to external considerations arising from quantum gravity to justify the introduction of a finite cut-off. But besides their speculative character and the desire to implement the original heuristics of the QFT program (as discussed below), it is unclear whether there is anything else to justify the “illegitimacy" of such appeal.

good reasons to believe that the predictions of $\mathcal{L}_{\text{SM}}$ will break down at some scale and to make this feature explicit in its formulation. Of course, as soon as we take into account more "speculative" considerations (e.g., the fate of the SM given quantum gravity), they obviously tip the scale in favor of the SMEFT. Yet, the important point is that even if we put these considerations aside, there are still comparatively good reasons to introduce a finite cut-off, i.e., the cut-off is not just introduced for the purpose of solving mathematical issues.

I have considered a relatively standard meaning of '*ad hoc*ness' so far: namely, the cut-off hypothesis is introduced for the sake of saving a model from refutation (assuming that a model should be rejected if it is inconsistent). There is yet another relevant meaning of '*ad hoc*ness' at play here. Philosophers and mathematical physicists eager to formulate an exact and mathematically rigorous formulation of QFT often take it to be a defect of effective QFTs that they do not exactly satisfy the "fundamental principles of relativistic quantum physics" (Buchholz, 2000, p. 2; see also D. Fraser, 2006, 2008, 2011). The worry, in other words, is that the introduction of a finite cut-off is ill-integrated with the principles of relativistic QFT and therefore somewhat contrived or unnatural given the original heuristics of the QFT program in high energy physics.[14] In general, a finite cut-off indeed breaks at least some of the core symmetries of a QFT at the level of its state space (e.g., the separation between high- and low-energy field configurations is not invariant under momentum translation, even with a smooth cut-off). And if we endorse stringent standards of mathematical rigor, the lattice formulation of QFTs fully breaks Lorentz invariance.

How should we respond to this new concern? The best answer, I believe, is to point out that there is no reason to expect the heuristics of a research program to be infallible.[15] The principles of quantum mechanics and special relativity may well fail to be exactly unified in a realistic

[14]For a discussion about this specific notion of *ad hoc*ness in the philosophy of science literature, which is also often associated with the idea of "coherence", see, e.g., Leplin (1975) and Schindler (2018).

[15]Another response is to point out that effective QFTs "approximately" satisfy the original principles of relativistic QFT—by emphasizing, for instance, that even lattice QFTs (and not merely their predictions) are "approximately" Lorentz invariant (cf. Williams, 2019b, pp. 17-18). One might be worried (in particular) about the ambiguous notion of "approximation" at work here, especially when it comes to comparing different theories or symmetry groups (e.g., is it "good enough" that a lattice QFT respecting some hypercubic symmetry takes the form of a Lorentz invariant continuum QFT in the zero lattice spacing limit?). I will leave this issue aside since it would require a far more extensive discussion than I can provide here.

continuum field theory and many historical cases support such cautious attitude, including the ultimate empirical irrelevance of the ether hypothesis after three hundred years of attempts to use it to develop models of gravitational and electromagnetic phenomena. We might also justify the negotiable character of these heuristics by defending a more pragmatic approach towards foundational matters—assuming, again, that what drives foundational inquiry is to understand what makes our most reliable models work in the first place. As Gross felicitously puts it,

> I am not sure it is necessary to formulate the foundations of QFT, or even to define precisely what it is. QFT is what quantum field theorists do. For a practicing high energy physicist, nature is a surer guide as to what quantum field theory is as well to what might supersede it, than is the consistency of its axioms. (1999b, p. 56)

The point is of course not that a foundationalist attitude towards QFT is necessarily mistaken, but rather that it is poorly motivated to reject the cut-off hypothesis simply because it does not fit with what we imagine QFT to be based on former research programs.

In a similar vein, we might also be worried that the finite cut-off does not arise “from within" compared to other natural scales. In classical electromagnetism, for instance, the speed of light arises as a direct consequence of combining Maxwell’s equations in a vacuum with one another (i.e., c is the inverse square root of the product of the permittivity and permeability of free space). By contrast, the high-energy cut-off in our best current effective QFTs is introduced “by hand" as a new dimensionful scale, and we might take its extrinsic and arbitrary character to provide one more reason to believe that its introduction is poorly motivated.

In response, we should first note that the imposition of a cut-off “by hand" is not as problematic as we might think. Existing cases of low-energy effective theories, such as the Fermi theory of beta decays, support the view that the cut-off is ultimately determined by the content of a more fundamental and complete theory. We do not yet have such theory for our best current effective QFTs. But it is not as if this should lead us to think that the introduction of a cut-off is absolutely ungrounded or as if there is no matter of fact about its exact value (or at least about some physically salient value). In QFT, the cut-off typically corresponds to the mass of a heavy field unaccounted for by the effective theory of interest and is thus naturally interpreted as a threshold energy for the production of a new type of heavy particle.

Still, the absence of empirical inputs at the relevant scales in current collider experiments implies that the value of the cut-off in our best effective QFTs is not fixed, and we might take this to be a good enough reason to reject the cut-off hypothesis. There is indeed something artificial or cooked-up about imposing an upper bound on the domain of a model if we do not yet have reliable means to determine its exact value. And to make the matter even worse, the probable breakdown of our best QFTs at high energies can be modeled with widely different types of cutoffs. As Wallace (2006, 2011) and Williams (2019b) have already emphasized, renormalization methods can be used to show that the low-energy content of such QFTs is typically largely independent of the exact value and type of the cut-off (setting aside the naturalness problem). Yet, we might still take it to be a defect of these QFTs that they contain such an arbitrary and somewhat idle feature (or “surplus structure").

I think that the final word on this specific issue is twofold. First, if we have good reasons to believe that some model is incomplete, it does not seem to be a default of the model that it contains an arbitrary scale parametrizing its probable predictive breakdown, regardless of whether it accurately represents the world or not. The introduction of an arbitrary cut-off in the formulation of our best models appears to be even the most reliable way to account for our ignorance about potentially new types of high-energy physics—the most “responsible" way to be ignorant, as it were. Second, if we have appropriate empirical inputs at low energies and endorse the naturalness assumption, i.e., the assumption that the dimensionless parameters of an effective theory are of order one, the value of the cut-off remains a prediction of the theory. The cut-off hypothesis thus fails to be *ad hoc* in its most traditional sense, i.e., it is falsifiable or empirically testable (see, e.g., Hesse, 1961; Popper, 1965). And even if the naturalness assumption fails in some cases, as it is currently the case with the (normalized) mass parameter of the Higgs field for instance, the determination of the value of the cut-off remains nonetheless a matter of experimental investigation.[16]

[16]See, e.g., Williams (2015) for a philosophical analysis of the naturalness problem.

But perhaps the “best", say, the most simple and elegant formulation of an incomplete or non-fundamental theory should not include any explicit restriction on its domain. As Baker puts it:

> Wallace rightly points out that QFT cannot be trusted to accurately describe reality outside its domain of application, in the very high-energy regimes which have not yet been experimentally investigated and where the laws of quantum gravity can be expected to apply. He sometimes suggests that this is a reason to prefer theories like LQFT [Lagrangian QFT] which break down at high energies (Wallace, 2011, p. 120). But the upshot is rather that whatever version of QFT we interpret, we should not trust its implications within the high-energy domain. By analogy, we trust Newtonian mechanics only in the low-speed domain where it approximates relativity, but this is no reason to prefer a theory that breaks down as objects approach the speed of light. Still, Wallace is right that when standard axioms for AQFT assign observables to every arbitrarily small region, this is not necessarily a virtue. (Baker, 2016, pp. 5-6)

Although Baker concedes that the extension of a theory (or model) beyond its regime of validity is not necessarily a good thing, he still suggests that there is something virtuous about working with the putatively complete version of a theory, even if we know it to be limited in some way or another. Baker does not explain why. But we can imagine at least two main reasons for it. First, the introduction of an explicit cut-off in a theory is interpretatively irrelevant if we can appropriately interpret the theory with the help of relevant pragmatic considerations concerning its domain of applicability: for instance, by keeping in mind that the theory is not meant to give a fundamental or complete picture of the world. Second, the theory without any explicit cut-off exhibits important theoretical virtues compared to its cut-off counterpart, such as simplicity and elegance for instance. We might take those virtues to counterbalance the fact that the structure of the theory is not explicitly adjusted according to its limited scope.

I will discuss the second point below. In response to the first, having a theory which contains explicit information about its probable empirical limitations does make the interpretative task less arbitrary. Without such information, we might simply disagree about the best interpretation of a theory depending on how seriously we take its probable future success, or we might take some parts of a theory more seriously than they deserve to be taken. As Williams (2019b) rightly argues in the context of QFT in my view, we are less likely to fall into those types of interpretative snares if we adjust the explanatory scope of a theory with the set of phenomena it accurately predicts and discard its explanatorily irrelevant parts (which is, more generally, the central message of the selective strategy advocated by Kitcher, 1993, Psillos, 1999, and Chakravartty, 2007, for instance).

The issue of *ad hoc*ness is not entirely settled yet. It seems that we can still modify at will the structure of an effective theory to make it fit with any sort of data. As we probe higher energy scales in experiments, we can add higher-order operators and adjust their parameters (i.e., the Wilsonian coefficients $C_k^{(i)}$ in Eq. 2.2) to compensate for experimental anomalies. In a way, an effective theory appears to have the ability to escape any kind of empirical refutation just as a polynomial equation specifying some curve can be adjusted at will to make it fit with a discrete set of data points (see, e.g., Forster and Sober, 1994; Sober, 2008, chap. 1, for a discussion of the curve-fitting problem). And so it seems that the predictions of an effective theory break down only insofar as it fails to be predictive, i.e., when we have been forced to introduce a sufficiently large number of operators to accommodate data and find ourselves without enough empirical inputs to fix the free parameters of additional operators (with the infinite expansion as a limiting case).

This last point, however, is not entirely correct. An effective theory is still defined by specifying a particular set of degrees of freedom and symmetries. These are, in turn, used to specify the types of phenomena that the theory can account for (e.g., the particles that we detect in experiments). And we may perfectly detect new short-lived particle tracks in high-energy collisions which are not predicted by the effective theory at stake. This means that an effective theory remains falsifiable despite having a highly flexible structure.

Likewise, the possibility of adding increasingly many "correction terms" in an effective theory does not make it *ad hoc* or cooked-up either. This is simply the wrong way of looking at the matter. As explained above, if we keep a finite cut-off and require the exact predictions of the model to be cut-off independent, we are forced to introduce all the possible terms consistent with the degrees of freedom and symmetries of the model. Equivalently, if we eliminate high-energy degrees of freedom in a given model, the resulting low-energy effective model reproducing exactly the predictions of the original model at low energies includes all the possible terms consistent with its symmetries. That is, when we formulate an effective model such as $\mathcal{L}_{\text{eff}}$, we are not saving some original model from refutation by adding correction terms and fine-tuning their parameters. If the predictions of the original model are likely to break down at some scale, the most complete and reliable description of the physics associated with the set of degrees of freedom and symmetries of the model in its limited domain includes all possible terms allowed by its structure. There is nothing principled about truncating the expansion of the model at a specific order as opposed to another (including the lowest orders). The appropriate choice of truncation depends on some desired accuracy and on the number of available empirical inputs. And so it appears to be even more *ad hoc* and arbitrary, albeit convenient, to restrict the model to a specific order than to include all the terms allowed by its structure.

2.4.2 Complexity

The case of complexity is more straightforward and overlaps in important ways with the case of *ad hoc*ness. I will leave aside the aesthetic notion of simplicity implicit in Redhead's remark above. The world could turn out to be severely tangled at its most fundamental levels and most beautifully depicted by laws that reflect its complexity. Concerning the second part of Redhead's remark, I also suspect that physicists would find it intellectually *unexciting* to learn that the project of physics has reached its completion and that there is nothing new to be discovered. Be that as it may, we might still believe that the apparent complexity of effective QFTs reflects their ability to accommodate data rather than to accurately represent unobservable entities or structures.

Philosophers of science traditionally identify and clarify what they take to be epistemically significant about the ambiguous notion of simplicity (or parsimony) by distinguishing between syntactic simplicity and ontological simplicity. A theory (or model) is syntactically simple if it contains a small number of basic principles while a theory is ontologically simple if it posits a small number of different kinds of entities (see, e.g., Schindler, 2018, chap. 1, for a recent overview of the debate over the criterion of simplicity). The good news is that, according to this classification, $\mathcal{L}_{\text{eff}}$ is as simple as $\mathcal{L}_{\text{SM}}$. The two models have the same particle content and the same core principles and symmetries. If we count perturbative renormalizability as one additional basic principle, $\mathcal{L}_{\text{eff}}$ even appears to be syntactically simpler. Agreed, $\mathcal{L}_{\text{SM}}$ has a higher degree of symmetry thanks to the existence of accidental symmetries. But overall, this does not seem to make $\mathcal{L}_{\text{eff}}$ more syntactically complex than $\mathcal{L}_{\text{SM}}$. And we could of course further restrict the form of $\mathcal{L}_{\text{eff}}$ by enforcing these accidental symmetries (cf. Fig. 2.1 above).

Still, the formal expression of $\mathcal{L}_{\text{eff}}$ is much more complex than that of $\mathcal{L}_{\text{SM}}$ insofar as $\mathcal{L}_{\text{eff}}$ involves an infinite sum of distinct operators. Does this affect in any way the ability of $\mathcal{L}_{\text{eff}}$ to be approximately true? It does not seem to be the case. We might re-express $\mathcal{L}_{\text{eff}}$ at will in a more concise form without affecting its content. For instance, we might express $\mathcal{L}_{\text{eff}}$ by means of a simple abstract sum over operators and hide, as it were, the formal complexity inherent in the exact expression of operators at each order. The real issue with $\mathcal{L}_{\text{eff}}$ rather lies in the fact that it contains an infinite number of independent parameters (including the cut-off). Continuing with the analogy of section 2.3, $\mathcal{L}_{\text{eff}}$ is similar to a mathematical table with an infinite number of independent entries or “inputs". To be sure, the structure of the table is governed by a small number of principles. But its size still suggests that it is meant to systematically organize data and save phenomena, and not to reveal anything special about unobservable entities or structures.

The appropriate response here is not to deny that $\mathcal{L}_{\text{eff}}$ displays a high degree of formal complexity but rather to examine whether it is justified. After all, we may have good reasons to prefer a theory with more free parameters than less in some contexts. And, as it turns out, the formal complexity of $\mathcal{L}_{\text{eff}}$ is a direct consequence of the cut-off hypothesis. Once we introduce a finite cut-off, the model automatically generates predictions containing an infinite number of cut-off dependent terms. As already emphasized, the only way to make exact predictions with the model which do not depend on the cut-off is to introduce all the operators allowed by its original symmetries. But this means that if we have any good reason to introduce a finite cut-off and derive exact predictions (or predictions as accurate as possible), we automatically have a good reason to introduce an infinite number of operators with arbitrary parameters (or as many as needed). Hence, the previous justifications for introducing a finite cut-off apply here too. Second, we need to specify some desired accuracy in order to restrict the set of operators in $\mathcal{L}_{\text{eff}}$. This sort of restriction is, of course, what makes $\mathcal{L}_{\text{eff}}$ predictive and computationally powerful in the first place. But it has nothing to do with what the world is like; the restriction only depends on our experimental and computational limitations. If we wish to reduce the formal complexity of an effective model in a principled manner, we need to elevate at least one additional pragmatic constraint as a basic principle of the model and thus increase its syntactic complexity. Perturbative renormalizability is, in this sense, just one such constraint among many others.

What should we make of the irreducible formal complexity of $\mathcal{L}_{\text{eff}}$ close to the cut-off scale? In this case, higher-order operators in $\mathcal{L}_{\text{eff}}$ cannot be ignored anymore, and it becomes quickly impracticable to compute predictions with the model. Here we need to keep in mind that effective theories are not meant to give a description of the world that works across all scales. Once a particular effective model becomes deficient close to some scale, we need to replace it by another model, perhaps by an extension of the effective model, or perhaps by something completely new. Either way, the important point is that the limitation of the effective model close to its cut-off scale is not a vice; it is even a virtue since the effective model signals from within, as it were, that we need to replace it by a more comprehensive one. Otherwise, well below the cut-off scale, the effective model provides a complete description of the system specified by a given set of degrees of freedom and symmetries, in the sense that the model can in principle take into account any kind of new physics that might affect the system within this limited domain in its own terms.

2.5 Conclusion: A Few Remarks About Gell-Mann's Totalitarian Principle

The Standard Model (SM) of particle physics is widely believed to be best formulated as an effective theory rather than as a putatively fundamental theory, and the probable limitations of the SM at short distances make this intuitively attractive. A major cause of concern, however, comes from the apparent *ad hoc*ness and complexity of the formulation of the SM as an effective theory. I have argued that these two vices are merely apparent. The introduction of a finite cut-off scale is both natural and justified insofar as we have good reasons to believe that the SM does not reliably apply across all scales, and we do have such reasons independently of its fate in light of future developments in physics. Moreover, once we introduce such a cut-off scale, the most complete and syntactically simple formulation of the theory includes all the terms allowed by its core principles. Any definitive restriction of the theory to a specific order, including the simplest one, requires us to elevate some pragmatic constraint as a basic principle of the theory and therefore makes its formulation all the more *ad hoc* and syntactically complex.

I would like to conclude with a few remarks about Gell-Mann's totalitarian principle, which states that everything that is not forbidden is compulsory.[17] As it turns out, effective theories perfectly embody this principle. By construction, an effective theory includes all the terms which are allowed by its core principles—anything less is not really the "full" effective theory and arguably not even an effective theory, strictly speaking. But there is some ambiguity about how to best interpret the totalitarian principle in this case. Interpreting this principle as a principle of consistency does not work since there is nothing inconsistent about dropping or adding higher-order terms in an effective Lagrangian (cf. Wells, 2012b, p. 70). Interpreting the totalitarian principle as a principle of sufficient reason does not work either (cf. Weinberg, 1999, p. 246). There *is* a reason for restricting an effective theory, say, to the second order as opposed to the third one. We might take this reason to be ultimately pragmatic, but it is still a reason. By the same token, interpreting the totalitarian principle as a principle of plenitude does not work (cf. Kane, 1976, p. 30; 1986, p. 130; Kragh, 1990, p. 272; 2019; Bangu, 2008, p. 246, footnote 26; Schulte, 2008, pp. 297, 310). The relevant version of the principle of plenitude here is that a term is to be included if there is no

[17]Gell-Mann briefly expressed this principle in the context of nuclear physics as a heuristic to identify the particle decays which are allowed and forbidden: "Among baryons, antibaryons, and mesons, any process which is not forbidden by a conservation law actually does take place with appreciable probability. We have made liberal and tacit use of this assumption, which is related to the state of affairs that is said to prevail in a perfect totalitarian state. Anything that is not compulsory is forbidden." (1956, p. 859, footnote (*)) Note that Gell-Mann uses the contrapositive of the principle in this last sentence and not the "converse" contrary to what Kragh (2019, p. 3) says, and so the content of what Kragh calls the "principle of compulsory strong interactions" is the same as the content of the totalitarian principle.

reason for it not to be included.

The best interpretation, in my view, is to treat the totalitarian principle in the specific case of effective theories as a principle of parsimony. The principle requires us to minimize the number of basic principles or constraints underlying the formulation of a theory. Once we endorse this principle, we do not need to introduce particular prescriptions each time we encounter unexpected situations within the limited domain of the theory. The simplest rule is to include all the possible terms in the first place and thus avail ourselves of the means to directly deal with any anomaly, even the most insignificant one.

Chapter 3: Effective Theories and Infinite Idealizations: A Challenge for Scientific Realism

Williams and J. D. Fraser have recently argued that effective field theory methods enable scientific realists to make more reliable ontological commitments in Quantum Field Theory (QFT) than those commonly made. In this chapter, I show that the interpretative relevance of these methods extends beyond the specific context of QFT by identifying common structural features shared by effective theories across physics. In particular, I argue that effective theories are best characterized by the fact that they contain intrinsic empirical limitations, and I extract from their structure one central interpretative constraint for making more reliable ontological commitments in different subfields of physics. While this is in principle good news, this constraint still raises a challenge for scientific realists in some contexts, and I bring the point home by focusing on Williams's and J. D. Fraser's defense of selective realism in QFT.

3.1 Introduction

There is a deeply entrenched strategy in philosophy of physics about how to interpret our best theories in realist terms. Philosophers usually start by pretending that the theory at stake is complete, true and final, even if it is known not to be true in all respects. Then, they eliminate its redundant parts by implementing sophisticated constraints on its structure. And eventually, they draw from the resulting theory some putatively complete picture of the world. The goal, ultimately, is to identify a definite set of unobservable entities or structures, whether they are actually fundamental or not, and thereby lay the ground for explaining the success of the theory in realist terms.[1]

[1]For a critical discussion of this traditional interpretative strategy, including references in the literature, see Ruetsche (2011, chap. 1) and Williams (2019b).

As it turns out, this strategy somewhat falls apart in the case of our most fundamental and empirically successful theories. We do not yet know whether realistic Quantum Field Theories (QFTs) can be consistently defined across all scales and therefore whether we can even consistently speculate about the possible worlds in which these theories are exactly true. Wallace (2006, esp. sec. 3.3; 2011), Williams (2019b), and J. D. Fraser (2018; 2020b) have proposed a more modest and cautious strategy in response, which is also better suited to the limited success of current and past theories. They enjoin philosophers to identify the ontological commitments necessary to explain the success of our best QFTs in the limited regimes where they are known to be reliable and not in the regimes where they are likely to break down.

The crucial part of Wallace, Williams and J. D. Fraser's proposal resides in the set of techniques they employ to implement this new strategy, namely, Effective Field Theory (EFT) methods (including the Wilsonian Renormalization Group). Broadly speaking, these methods have been developed in QFT to treat phenomena at different scales separately, and they became popular in physics in large part because of their remarkable heuristic, computational and predictive power. More crucially for interpreters, the QFTs constructed by using these methods, i.e., EFTs, are intrinsically restricted to some limited range of distance scales. The physics within this range can even be shown in typical cases to be largely independent of the specific details of the short-distance physics. And this has led Williams and J. D. Fraser, in particular, to argue that EFTs provide a more perspicuous and reliable interpretative standpoint to identify unobservable entities or structures in the appropriate regimes, even if realistic QFTs are ultimately shown to be consistent across all scales.

This chapter has two closely related aims. The first is to show that the interpretative relevance of EFT methods extends beyond the specific context of QFT. Given that most if not all known physical systems exhibit distinct scales in most circumstances, it should come as no surprise that the EFT paradigm has been successfully implemented in most areas of contemporary physics during the last decades.[2] Yet, we might still wonder whether the theories constructed by using EFT methods share

[2]For references to the extension of EFT methods outside condensed matter and particle physics, see, e.g., Endlich et al. (2011), Dubovsky et al. (2012), and Gripaios and Sutherland (2015) for fluid dynamics; Donoghue (1995)

distinctive structural features that might help us make more reliable ontological commitments in different subfields of physics. I will first argue that effective theories are best characterized in general by the fact that they contain intrinsic empirical limitations, i.e., their structure incorporates a robust specification of the scales at which they are likely to be empirically inaccurate before we probe these scales in experiments. This contrasts with the usual situation where the empirical limitations of a theory are found only by a direct confrontation with experimental data obtained at the relevant scale. Then, I will briefly present and justify the realist account of effective theories which follows the most naturally from this characterization. I will call it the "Standard Effective Account" and show that the structure of an effective theory forces us to restrict our commitments to entities or structures which can be specified within the limited range where the theory is likely to remain empirically reliable.

The second aim is to assess whether Wallace, Williams and J. D. Fraser's strategy enables scientific realists to fulfill their explanatory duties. Starting with the traditional form of scientific realism (cf. Psillos, 1999, pp. xvii-xix), I will first give a concrete example of the restrictions we face if we treat our best current theories as effective theories.[3] We may think, for instance, that we have good reasons to take the descriptions of continuum fields in the effective versions of the Standard Model of particle physics and General Relativity to be approximately true and therefore to commit to the existence of those entities, i.e., of continuous systems with an infinite number of degrees of freedom. I will argue that on the Standard Effective Account, we cannot reliably make such ontological commitments. And my point here is not so much to claim that infinite physical systems are beyond our ken—in a way, we have known this for a long time—but rather to

and Burgess (2004) for general relativity; Goldberger and Rothstein (2006) and Porto (2016) for post-Newtonian gravitation; Baumann and McAllister (2015, chap. 2) and Burgess (2017) for inflationary cosmology; Polchinski and Strominger (1991) and Hellerman et al. (2014) for low-energy string theories; Baumann and Green (2012) and Kaplan (2016, esp. sec. 8.4.3) for advanced topics relevant to quantum gravity.

[3] Of course, this requires assuming that we do not yet have some decisive evidence that we have hit a true, final and complete theory in physics or some complete theory providing an approximately true description of the world in all respects. We also need to assume that effective theories display sufficiently many theoretical virtues to be even considered candidates for making approximately true claims about the world (see Wells, 2012b, chap. 5, for a discussion related to this point). We do not need, however, to deny the existence of a final theory, which is implicit in the traditional scenario of an infinite "tower" of EFTs, where each theory of an endless series of EFTs describes phenomena within a limited range of energy scales.

illustrate how the structure of effective theories imposes clear-cut restrictions on one's ontological commitments.

I will then argue that, in some specific theoretical contexts including classical and quantum field theory, these restrictions still raise a challenge for more refined forms of scientific realism. To bring the point home, I will focus on Williams's (2019b) and J. D. Fraser's (2018; 2020b) defense of selective realism in QFT and, expanding on Ruetsche's (2018; 2020) discussion, show that the candidates which look at first sight the most appealing for making ontological commitments in the appropriate regimes—namely, correlations, particles, and lattice fields—fail in other important respects. The best candidates that do not suffer from the same issues appear to be continuum fields, with the proviso that they are approximately similar to large distance scale features of the world. But, again, selective realists cannot take the descriptions of continuum fields to be approximately true *simpliciter*, which leaves them with no obvious candidate for offering a genuine defense of the realist cause. I will conclude briefly with a more radical suggestion to circumvent this issue: namely, to modify the standard semantic tenet of scientific realism endorsed by selective realists (e.g., Psillos, 1999; Chakravartty, 2007) and index (approximate) truth to physical scales.

The chapter is organized as follows. Section 2 presents two distinct examples of effective theories. Section 3 argues on the basis of these examples that effective theories are best characterized by the fact that they contain intrinsic empirical limitations. Section 4 presents the Standard Effective Account. Section 5 shows that traditional scientific realists cannot commit to the existence of the infinite systems specified by a literal interpretation of our best effective theories. Section 6 extends the discussion to Williams's and J. D. Fraser's defense of selective realism.

3.2 Two Examples of Effective Theories

Philosophers have not paid much attention to the diversity of effective theories across physics (e.g., Cao and Schweber, 1993; Hartmann, 2001; Bain, 2013); and when they treat the particular case of EFTs in particle and condensed matter physics as a new paradigm for understanding physical theories, they often remain too elusive or attribute too much importance to parochial features absent in other types of effective theories. For instance, it is common to characterize effective theories as theories that directly incorporate into their mathematical structure the imprint of their breakdown at some non-trivial finite physical scale (e.g., Bain, 2013, p. 1; Williams, 2019a, p. 2; 2019b, pp. 6-7, 9–10, 13). But seldom is it specified whether, in the general case, effective theories display some mathematical singularity, become physically meaningless, make inconsistent predictions, or become merely empirically inaccurate at that scale.[4] In order to give a sufficiently comprehensive and informative characterization, I will thus first present two different kinds of effective theories and examine, in particular, the way in which they “break down" at some scale.[5]

Example 1: Consider first the mathematically most simple formulation of the Newtonian gravitational theory for a body of mass m_1 interacting with another body of mass m_2:

$$m_1 \frac{d^2 r}{dt^2} = -m_1 \frac{m_2 G}{r^2} \tag{3.1}$$

with r the relative distance between the centers of mass of the two bodies and G the Gravitational constant.

There are two distinct ways to construct an effective version of this theory. Since we already know its closest successor, i.e., classical General Relativity, we can simply follow the “top-down" strategy: namely, we appropriately restrict the range of parameters of the more comprehensive theory and eliminate its theoretical constituents which do not contribute significantly to predictions within this range. For instance, we can derive Eq. 3.1 with additional correction terms encoding relativistic effects by implementing weak-gravity and low-velocity restrictions on the simplest solutions to the equations of classical General Relativity (see, e.g., Poisson and Will, 2014, for more details).

We can also pretend that we do not yet know the more comprehensive theory and follow the “bottom-up" strategy. We first identify a limited range where we think that the theory provides

[4]Other overly broad characterizations include “approximate theories" (e.g., Castellani, 2002, p. 263; Ruetsche, 2020, p. 7), “non-fundamental theories" (e.g., Egg, Lam, and Oldofredi, 2017, p. 455), and “phenomenological theories" (e.g., Huggett and Weingard, 1995, p. 189; Butterfield and Bouatta, 2014, p. 65).

[5]For simplicity, I will understand ‘theory’ in its specific sense throughout the chapter, that is to say, as given by a specific action, a Lagrangian or a Hamiltonian—or even more simply by equations of motion.

reliable information. For instance, we may suspect from the infinite value of $m_1 m_2 G/r^2$ in the limit $r \to 0$ that Eq. 3.1 becomes mathematically inadequate for describing the gravitational interaction between arbitrarily small bodies moving arbitrarily close to one another. Or we may have already found that the theory makes slightly inaccurate predictions when the gravitational force $m_1 m_2 G/r^2$ becomes too strong. Then, we restrict the range of the theory by introducing some arbitrary limiting scale, namely, a short-distance scale r_0 in this case. And finally, we include all the possible terms depending on r_0/r which are allowed by the symmetries of the theory, with one arbitrary coefficient for each new term. As we perform these steps, we do not need to know anything about the value or the underlying meaning of the limiting scale, namely, that r_0 turns out to be the Schwarzschild radius $2m_2 G/c^2$ of the body of mass m_2, with c the speed of light. The value of the additional coefficients and r_0 is ultimately determined by means of experimental inputs, at least for a finite number of them.[6]

Now, whether we follow the top-down or the bottom-up strategy, the resulting effective theory takes the following form:

$$m_1 \frac{d^2 r}{dt^2} = -m_1 \frac{m_2 G}{r^2}\left(1 + a_1 \frac{r_0}{r} + a_2 \left(\frac{r_0}{r}\right)^2 + a_3 \left(\frac{r_0}{r}\right)^3 + ...\right) \tag{3.2}$$

with a_1, a_2, a_3, etc. some arbitrary coefficients. The most complete version of Eq. 3.2 includes an infinite number of terms which depend on r_0/r and leave the equation invariant under Galilean symmetry transformations (i.e., translations in space and time, spatial rotations, and velocity boosts). We can also define an effective theory by means of a finite number of terms and fix the value of their coefficients by means of experiments.[7]

How should we interpret the scale r_0 if we take the structure of these effective theories at face value? Suppose for the sake of the argument that we are interested in predicting the value of the acceleration d^2r/dt^2 in Eq. 3.2. The first thing to note is that the contributions of higher-

[6] In general, we also need to assume that the dimensionless constants of the theory are of order 1 to estimate the value of the limiting scale, i.e., we need to endorse the "naturalness" principle $a_i = O(1)$ in Eq. 3.2 below.

[7] For more details about the first-order relativistic and quantum corrections to the non-relativistic gravitational potential, see, e.g., Donoghue (1995), Burgess (2004), and Blanchet (2014). Note that, in some cases, existing empirical measurements (or some other reason) may require us to break some of the symmetries of the original equation.

order terms $(r_0/r)^n$ to predictions are negligible for $r \gg r_0$ and very large for $r \ll r_0$. If we include increasingly many higher-order terms in Eq. 3.2, the predictions remain overall the same for $r \gg r_0$ and become increasingly large around and below r_0. And if we include an infinite number of terms, the resulting expansion $\sum_i a_i(r_0/r)^i$ takes an infinite value for $r_0/r \geq 1$. Hence, if we simply look at the mathematical structure of the family of effective theories associated with Eq. 3.2, we find that their predictions display a sharp pattern of variation around the characteristic scale r_0, which remains robust as we add or remove higher-order terms.

At first sight, this predictive pattern does not appear to tell us much about r_0 since the expansions $\sum_i^N a_i(r_0/r)^i$ for finite N are mathematically well-defined across all distance scales (except for the trivial scale $r = 0$). Yet, if we consider these finite expansions in relation to one another, we learn that we can always add small correction terms of increasing order in r_0/r in any given expansion if we want to improve its predictive accuracy for $r \gg r_0$. And if we consider these finite expansions in relation to the limiting case of the infinite expansion, we also learn that they ultimately become mathematically ill-defined at r_0 when we add increasingly many such terms. In short, if we try to make any of these finite expansions as predictively accurate as possible for $r \gg r_0$, we end up with theories making infinite predictions at r_0 and below, i.e., with theories which, as a matter of principle, cannot be empirically accurate for $0 < r \leq r_0$. And this, in turn, provides at least preliminary reasons to believe that the pattern of variation around r_0 does not simply reflect some notable qualitative physical change but rather signals that these finite expansions are likely to become unreliable around r_0.

Now, this interpretation is grounded in the experimental profile of existing theories displaying the same predictive pattern. If, for simplicity, we use Eq. 3.2 as an example, the experimental pattern takes the following form. We start with some effective theory defined by means of a finite expansion and fix its parameters by means of experiments at large distance scales r. At shorter distance scales, however, we find small experimental discrepancies and decide to add new terms to compensate for them. Yet, as we probe even shorter distance scales, the effective theory with the additional terms becomes all the more quickly empirically inaccurate and we need, at least in principle, to introduce new terms if we want to maintain its predictive power and accuracy. In practice, physicists directly look for a new theory in situations like this. If we were to keep up with the original theory and probe phenomena closer and closer to r_0, however, we would need to introduce an infinite number of terms. Since all these terms are equally important at r_0, we would not be able to select a finite number of them in order to make approximate predictions. And since we cannot in practice make an infinite number of measurements to fix the value of an infinite number of arbitrary coefficients, the theory would lose its predictive power. Hence, according to this pattern, r_0 corresponds to the maximal predictive limit of the family of effective theories associated with Eq. 3.2. For the infinite expansion, r_0 corresponds both to a characteristic scale where the theory becomes mathematically ill-defined and predictively powerless. For the finite expansions, the demarcation is not as vivid and sharp; but, overall, the corresponding effective theories make empirically accurate predictions for $r \gg r_0$ and empirically inaccurate ones for $r \ll r_0$.

Note that the same argument does not apply to the original Newtonian theory in Eq. 3.1 despite its divergent behavior at $r = 0$. If we leave aside the apparent physical impossibility of the situation characterized by $r = 0$, we still face the issue that the limiting scale $r = 0$ is experimentally trivial from the perspective of classical Newtonian gravitation. Even if we can, in principle, probe the system down to arbitrarily short distances in this context, we can only gain experimental information about finite size effects resulting from the gravitational interaction between two bodies at some finite distance from one another. In the case of effective theories, the situation is different because there is no physical principle or experimental constraint which indicates that the regime specified by $r \leq r_0$ is either experimentally inaccessible or trivial. The infinite expansion becomes deficient at r_0. But nothing in the theory suggests that we cannot use bodies to probe distance scales within $0 < r \leq r_0$ compared, say, to string theory where we cannot use strings in scattering processes to probe distances shorter than the string scale (see, e.g., Hossenfelder, 2013, sec. 3.2, for a discussion).[8]

[8]Note, moreover, that we cannot define some non-trivial limiting distance scale r_0 by using only m_1, m_2, and G in Eq. 3.1. By dimensional analysis, we would need to introduce a new arbitrary velocity scale c and therefore modify

Example 2: Consider now a standard example of QFT, the ϕ^4-theory. The theory describes a simple quantum field, i.e., a continuum of smoothly coupled individual quantum systems over space-time with each system characterized by only one degree of freedom. In a somewhat analogous way as in Eq. 3.1, the original dynamical equation is given by:

$$\partial_\mu \partial^\mu \phi(x) + m^2 \phi(x) = -\lambda \phi^3(x) \tag{3.3}$$

where $\phi(x)$ is a real-valued variable describing a possible configuration of the field over space-time, ∂_μ the analog of d/dt in the four-dimensional Minkowski space-time, m a mass parameter, and λ a self-coupling parameter. This equation contains no explicit intrinsic limitation, which suggests that there is *a priori* no reason to believe that the theory fails to apply at arbitrarily large and short distances (or, equivalently, at arbitrarily low and high energies). The trouble comes when we try to compute predictions. Typically, in QFT, this is done by evaluating the correlations between some initial and final field configuration states characterizing some scattering process, where these states describe, roughly speaking, the particles that we prepare and detect in experiments. Calculating these correlations requires, in turn, including the contributions from all the possible transitions between these states and therefore summing over all the possible intermediary field configuration states. If we do that, however, the high-energy configurations of the field, i.e., the configurations which vary rapidly over short-distance scales, give rise to infinite probabilistic predictions, which is inconsistent.

As of today, the only way to solve this problem in realistic QFTs is to modify the structure of the theory by means of “renormalization" methods.[9] In the case of the ϕ^4-theory, for instance, we can smoothly lower the contributions of the high-energy field configurations $\tilde{\phi}(k)$ over some high-energy cut-off Λ by using a new field variable $\phi_\Lambda(x)$ with exponentially decreasing contributions above Λ:

$$\phi_\Lambda(x) \propto \int d^4k e^{ikx} (e^{-k/\Lambda} \tilde{\phi}(k)) \tag{3.4}$$

the structure of the original theory.

[9] See, e.g., Butterfield and Bouatta (2015), Williams (2019a), and Rivat (2019) for introductory discussions.

Similarly to Example 1, the value of the limiting scale Λ is not fixed at this stage. Yet, the QFT case is special. If we keep a finite cut-off, we can make the predictions of the theory Λ-independent by absorbing Λ-dependent terms into its parameters, at least for a finite range of values of Λ. But this requires including all the possible interaction terms allowed by the symmetries of the theory:

$$\partial_\mu \partial^\mu \phi_\Lambda(x) + m^2(\Lambda)\phi_\Lambda(x) = -\lambda(\Lambda)\phi_\Lambda^3(x) - g_5(\Lambda)\phi_\Lambda^5(x) - g_7(\Lambda)\phi_\Lambda^7(x) - ... \tag{3.5}$$

where the g_i's are new arbitrary coupling parameters depending on Λ. If we have appropriate experimental inputs, we can define an effective theory by means of a finite number of interaction terms, fix their parameters, and estimate the value of Λ (as in Example 1).

The predictive pattern in this example is overall similar to the one displayed in the previous example. Once we fix the parameters of the theory, we can show that the higher-order interaction terms $g_i(\Lambda)\phi_\Lambda^i$ in Eq. 3.5 contribute to predictions by increasing powers of (E/Λ), with E the characteristic energy scale of the scattering process considered. Yet, there is one crucial difference: the predictions of the theory typically become inconsistent for energies E close to and above Λ whether we include a finite or an infinite number of interaction terms in Eq. 3.5. Hence, if we take the structure of effective QFTs at face value, Λ is naturally interpreted as the scale at which the theory is likely to make inconsistent and *a fortiori* empirically inaccurate predictions.[10]

3.3 What is an Effective Theory?

Now that we are equipped with two different examples, let us look at several options for characterizing what is so distinctive about effective theories. I will argue that the structure of an effective theory is best characterized by the fact that it incorporates a robust specification of the scales at which it is likely to be empirically inaccurate (assuming, in particular, that we have appropriate experimental inputs to fix its free parameters).

Characterization 1: A first option is to characterize an effective theory as a low-energy limit

[10]As it turns out, the ϕ^4-theory is even more special: the perturbatively renormalized coupling $\lambda(\Lambda)$ diverges at some finite high-energy scale, i.e., it displays a “Landau pole" singularity.

of a more complete theory—even if this more complete theory is not fully known, which means that an effective theory is a particular realization of a given theory over a restricted range of energy scales. This relational characterization fits well with high-energy physicists' general description of EFTs (e.g., Burgess and Moore, 2006, p. xi, p. 456) and with the top-down Wilsonian procedure for deriving an EFT by eliminating high-energy field configurations.

To give a concrete example, suppose that the ϕ^4-theory is a low-energy realization of a more complete theory including a light scalar field $\phi(x)$ of mass m and a heavy scalar field $\psi(x)$ of mass M, with $m \ll M$. We can derive effective theories as follows. First, we eliminate, or "integrate out", the heavy field variable $\psi(x)$ in the high-energy theory (or, more precisely, in its functional path integral $\mathcal{Z}$). This gives rise to exotic terms depending on the variable $\phi(x)$ such as $\phi(x)(-\partial_\mu\partial^\mu + M^2)^{-1}\phi(x)$. Assuming that the characteristic energy E of the scattering processes of interest is much smaller than the mass of the heavy field, i.e., $E \ll M$, we can expand these exotic terms into an infinite series of polynomial terms depending only on the variable $\phi(x)$, its derivatives, and some inverse power of M. Schematically,

$$\begin{aligned} \mathcal{Z} &= \int \mathcal{D}[\phi]\mathcal{D}[\psi] e^{i\int d^4x \left[\frac{1}{2}(\partial_\mu\phi)^2 - \frac{m^2}{2}\phi^2 - \frac{\lambda}{4!}\phi^4 + \frac{1}{2}(\partial_\mu\psi)^2 - \frac{M^2}{2}\psi^2 - \frac{g}{4}\phi^2\psi^2\right]} \\ \Longrightarrow \mathcal{Z} &= \int \mathcal{D}[\phi] e^{i\int d^4x \left[\frac{1}{2}(\partial_\mu\phi)^2 - \frac{m^2}{2}\phi^2 - \frac{\lambda}{4!}\phi^4 - \frac{g_6}{M^2}\phi^6 - \frac{g_8}{M^4}\phi^8 - \dots\right]} \end{aligned} \tag{3.6}$$

with the appropriate coupling parameters g and g_i.[11] The structure of the effective theory is fully specified by the restrictions imposed on the high-energy theory with the appropriate low-energy assumption. In particular, since the contributions of the interaction terms $(g_i/M^{i-4})\phi^i$ give rise to inconsistent predictions close to M, the high-energy theory provides a natural high-energy cut-off for renormalizing the effective theory, namely, the mass of the heavy field. We can also define effective theories by restricting the series to some finite order in $1/M$ and obtain the original ϕ^4-theory by taking the limit $M \to \infty$.

The main problem with Characterization 1 is that it is either too broad or too narrow depending on how we understand it. If we take it to apply to any theory which is, in principle, derivable

[11] See, e.g., Baumann and McAllister (2015, sec. 2.1.1) and Petrov and Blechman (2016, sec. 4.1) for more details.

from a more complete theory in its low-energy limit, even indirectly, we may have reasons to suspect that it applies to all empirically successful theories built up so far. However, if we do not specify the structure of the high-energy theory or provide the specific details of the derivation, we will be left with a characterization which is overly vague and which, in particular, does not help us to circumscribe specific structural features common to effective theories. And to make the matter even worse, some standard cases of EFTs do not seem to have any high-energy completion and therefore to be even derivable, as a matter of principle, from a high-energy theory (see, e.g., Adams et al., 2006, for a discussion).

Inversely, if we take this characterization to apply only to theories which are explicitly related to a more comprehensive theory by means of some energy parameter or mass scale, as in Eq. 3.6, we will leave out many standard cases of effective theories, including Example 1. In general, the types of limiting scales and power counting schemes underlying the structure of effective theories, i.e., the rules for evaluating how the contributions to predictions of the different parts of an effective theory vary with some parameter, can be extremely diverse. Examples 1-2 illustrate this variety of scales. Example 1 provides a simple velocity power counting scheme when applied to a system of two bodies with the same mass m_1 and orbital radius r. In the non-relativistic regime, the virial theorem holds ($v^2 \sim Gm_1/r \sim r_0/r$), which means that the interaction terms in Eq. 3.2 contribute to predictions by increasing powers of the characteristic velocity v of the system. And it is more appropriate in this case to speak of a low-velocity realization of a more complete theory.

Characterization 2: A more promising strategy might be to look for some abstract feature internal to the mathematical structure of an effective theory.[12] Suppose for instance that we take an effective theory to be a theory which, while remaining mathematically well-defined over some limited range of parameters, becomes ill-defined at some non-trivial finite scale. This characterization fits well with the most complete versions of the effective theories presented in Examples 1-2 (e.g.,

[12]Appealing to a particular mathematical structure does not seem to give an adequate trade-off between generality and informativeness. The closest we can probably get to Examples 1-2 and standard cases of effective theories is to characterize the structure of effective theories in terms of Taylor (or Laurent) series in some parameter (or truncations thereof). But even then, this solution excludes exotic cases of effective theories with non-polynomial interaction terms in the field variables (see, e.g., Gripaios, 2015, sec. 5, for some models including such terms).

$\sum_i a_i(r_0/r)^i$). It also fits well with the attitude sometimes expressed in the philosophical literature according to which the framework of EFTs provides a general, efficient, and "opportunistic" way of solving the mathematical issues of QFTs (see, e.g., Butterfield, 2014, sec. V.2.2; Butterfield and Bouatta, 2015, sec. 3.1.3). Indeed, the very idea of introducing and keeping a finite cut-off is vindicated by the pathological behavior of QFTs at high energies (cf. Example 2). And even if we attempt to cure QFTs of their mathematical difficulties with renormalization methods, some paradigmatic cases like the ϕ^4-theory and Quantum Electrodynamics (QED), the quantum theory of the electromagnetic force, are likely to remain mathematically ill-defined at some large yet finite energy, i.e., to display a Landau pole singularity. If we want to define these pathological cases of QFTs consistently, they leave us with no choice but to restrict their range of parameters, and this suggests that EFT methods were meant to be applied to these sorts of theories.

Once again, however, this characterization excludes simple cases of effective theories and therefore appears to be too restrictive. For instance, the effective theories defined by means of a finite number of terms in Example 1 remain mathematically well-defined across all distance scales (except at the trivial scale $r = 0$) and therefore do not fall under Characterization 2. Agreed, being mathematically ill-defined at some non-trivial finite scale is presumably a sufficient condition for a theory to be characterized as effective (provided we introduce some cut-off); but these simple examples of classical point-particle effective theories show that this condition is not necessary.

Characterization 3: A third option, the one I favor, is to characterize effective theories by the fact that they contain intrinsic empirical limitations. Namely: an effective theory incorporates into its structure a robust specification of the ranges of scales where it is likely to be empirically inaccurate. There are four essential ingredients here:

1. The mathematical structure of the theory contains some non-trivial finite scale ("intrinsic limiting scale" or "cut-off");

2. It is possible to include increasingly many terms depending on this limiting scale which are consistent with the core principles governing the structure of the theory, with one arbitrary coefficient for each new term introduced;

3. These terms are systematically organized according to the importance of their contributions to predictions below and above the limiting scale ("power counting scheme");

4. As we include increasingly many such terms, the predictions derived from the theory remain approximately the same, say, below the limiting scale and become increasingly large around and above this scale ("robustness").

The predictive pattern is well illustrated by Examples 1-2, although it does not essentially depend on the particular details of their mathematical formulation, and, in general, the interpretation in terms of intrinsic empirical limitations is grounded in the experimental profile of existing theories displaying the same predictive pattern. Note as well that Characterization 3 does not imply that the mathematical structure of an effective theory delineates *by itself* the scales at which its predictions are likely to break down. We usually need to have experimental inputs in some accessible regime and assume that the dimensionless constants of the theory are of order one if we want to estimate the value of the limiting scale. Similarly, adding a list of provisos of the form 'For velocities much smaller than the speed of light' or '$r \gg r_0$' in the preamble of the theory is not sufficient: Characterization 3 requires the theory to have the imprint of its probable predictive failure directly written into its mathematical structure in the sense specified above.

Now, the advantage of this option is twofold. First, Characterization 3 is neither too restrictive nor too permissive. In particular, it applies to Examples 1-2 and standard cases of classical and quantum effective theories. It also excludes standard cases of theories putatively applicable across all scales such as the Newtonian theory defined in Eq. 3.1 and the perturbatively renormalizable version of Quantum Chromodynamics (QCD), the quantum theory of the strong force.[13] As explained in section 3.2, if we take such theories at face value, their structure does not explicitly delineate non-trivial experimental regimes where their predictions are likely to break down. Of course, we may impose a finite cut-off on the perturbatively renormalizable version of QCD because we suspect that QCD is likely to be empirically inaccurate at very high energies, and include

[13]This supposes that we set aside potential trouble at low energies and assume that the theory is sufficiently mathematically well-defined at arbitrarily high energies.

higher-order interaction terms into the theory. We may also exploit the hierarchy of scales exhibited by the different masses of the quarks in QCD and define a low-energy theory of the light quarks u, d and s with some cut-off because we suspect that it is easier to compute low-energy predictions if we eliminate irrelevant high-energy degrees of freedom. In both cases, however, we will be dealing with a different kind of theory, strictly speaking: namely, an effective theory which falls under Characterization 3.

Second, the characterization is also informative. Most remarkably, it offers a sharp distinction between two kinds of theories (or models): (i) theories with intrinsic empirical limitations, i.e., which already contain in their structure information about where they are likely to make inaccurate predictions before we probe the relevant scales in experiments; and (ii) theories with extrinsic empirical limitations, i.e., which are found to make inaccurate predictions only by a direct confrontation with experimental data obtained at the relevant scale. As we will see in the next section, the structure of an effective theory also gives good reasons to believe that it provides reliable ontological guidance only within a limited part of the world.

3.4 The Standard Effective Account

So far, I have argued that effective theories are best characterized by the fact that they contain intrinsic empirical limitations, but I have not said anything yet about their representational achievements. Suppose then that some effective theory is found to make accurate predictions within some regime and that its predictions are likely to break down at some scale beyond this regime. The most straightforward realist explanation in this case is to take the theory to accurately represent a limited part of the world and misrepresent, or fail to represent, other parts. Since this explanation fits well with the set of commitments shared by philosophers who explicitly defend a realist interpretation of EFTs, I will be relatively brief in this section. I will clarify the idea that the domain of applicability of effective theories is intrinsically limited by means of four common claims made about EFTs, briefly justify them by relying on general features of effective theories, call the resulting account the "Standard Effective Account", and extract one central interpretative constraint from it.

This is, of course, not to say that these philosophers agree on everything. There are indeed substantive interpretative disagreements in the literature on EFTs. But I will ignore those differences and restrict myself to extending the four common claims beyond the context of QFT.

The first difficulty here is that the term 'domain of applicability' is ambiguous. We could arguably take it to refer to the universe of discourse or interpretation of the theory, to the set of phenomena accounted for by the theory, to the range of variables specifying the possible physical states of the system described by the theory, or perhaps even to the range over which the theory is mathematically well-defined. If we keep in mind that the target of the theory is the actual world, the following notions should be sufficiently neutral and adequate for clarifying the Standard Effective Account. (i) The "domain of applicability" of a theory is the set of concrete physical objects—entities, structures, properties, quantities, states, phenomena, dispositions, and so on—that the theory accurately represents. The domain of applicability of a theory is not necessarily identical to its putative domain of applicability, i.e., to the set of putative physical objects specified by a literal interpretation of the theory.[14] (ii) The "domain of empirical validity" of a theory is the range of physical parameters over which its predictions are likely to remain accurate. If we have good reasons to believe that we have found a final theory, this domain ranges over all physically possible scales. Otherwise, if we do not have any means to estimate the empirical limitations of the theory in advance as in the case of effective theories or any evidence that the theory will remain empirically accurate in new regimes, this domain reduces to the range over which the theory has been found to be empirically accurate.

Then, the Standard Effective Account can be spelled out in terms of the four following claims:

1. The domain restriction claim: The domain of applicability of an effective theory is restricted by the limits of its domain of empirical validity (cf., e.g., Cao and Schweber, 1993, p. 76; Castellani, 2002, p. 260; Wallace, 2006, sec. 3.2-.3.3; Schweber, 2015, p. 60; J. D. Fraser 2018, p. 1173; Williams, 2019b, p. 13).

[14] By 'literal' I mean that the physically meaningful descriptions of the theory are understood in their standard sense and taken to be either true or false.

To take the simplest case of physical object, the domain restriction claim states that an effective theory accurately represents some concrete entity only if its core properties can be specified within the limited range where the theory is likely to remain empirically accurate. By 'core property' I mean that the property is constitutive of the identity of the entity (e.g., an infinite number of degrees of freedom for a continuum field). Now recall that if we have appropriate experimental inputs, say, at large distances, we can estimate the value of the limiting scale of an effective theory, say, a short-distance cut-off scale. And even if we have not yet probed phenomena close to this scale in experiments, the structure of the theory already gives us good reasons to believe that its predictions are inaccurate beyond this scale. As a realist, it is standard to assume that if a theory accurately represents the entities characterizing a specific domain, it also makes accurate predictions in this domain. Hence, the standard realist explanation of the probable predictive failure of an effective theory beyond its limiting scale is that the theory is likely to misrepresent, or fail to represent, the entities characterizing the corresponding domain (assuming here that there are such entities). And this means that the structure of an effective theory prevents us from simply remaining agnostic about its putative representational success beyond its limiting scale. We also have good reasons to think that the theory provides unreliable information about physical properties beyond this scale and therefore fails to give an accurate picture of the entities which are individuated by such properties.

In Example 2, for instance, the imposition of the smooth cut-off in Eq. 3.4 does not eliminate any degree of freedom in the original theory. On the face of it, then, the effective theory represents a putative continuum field with one degree of freedom at every point of space-time and therefore attributes core properties to its target system within any arbitrarily small region of space-time. At the same time, the pathological predictions of the theory around and beyond Λ also give very good reasons to believe that the theory misrepresents the structure of matter at arbitrarily short distances and therefore that it does not accurately represent a putative continuum field, strictly speaking. According to the domain restriction claim, however, it is perfectly possible for the theory to accurately represent, say, a real physical pattern of characteristic size larger than $1/\Lambda$

(see sections 3.5-3.6 for a discussion).

2. The new physics claim: The structure of an effective theory strongly suggests that the theory misrepresents or fails to represent some putative physical objects (cf., e.g., Robinson, 1992, p. 394; Cao and Schweber, 1993, p. 76; Wallace, 2006, sec. 3.2-.3.3; J. D. Fraser 2018, p. 1173; Williams, 2019b).

This claim is best supported by examining the relation between successive effective theories, or even the relation between an effective theory and some putatively fundamental theory. If we take effective theories in isolation, however, we can still give some support to this claim by relying on their structure. Consider Example 2 again. The effective version of the ϕ^4-theory with a smooth cut-off is mathematically well-defined at any point of space-time (at least according to physicists' standards) and does not contain any physical principle or constraint implying that the range beyond Λ is physically forbidden. To take again the simplest case of physical object, the theory thus appears to allow for the existence of concrete entities at arbitrarily short distances. Yet, as already emphasized, the theory also makes inconsistent predictions beyond Λ. Taken together, these two features strongly suggest that the theory is deficient in some way or another rather than that the world contains some physical limit at the scale Λ. And the best realist explanation, in this case, is that the theory does not include the appropriate theoretical constituents which would give rise to consistent predictions at short-distance scales and therefore that the theory either misrepresents or fails to represent putative entities at these scales instead of specifying, say, the fundamental graininess of space-time.[15]

3. The approximate truth claim: Effective theories offer approximately accurate representations in their domain of empirical validity (cf., e.g., Castellani, 2002, p. 260; J. D. Fraser 2018, p. 1173; Williams, 2019b, sec. 3).

[15] Note that the scale at which the predictions of an effective theory break down does not need to be *exactly* the same as the scale at which the new physics kicks in. For a discussion about the intricate link between violations of perturbative unitarity and the onset of new physics in the context of QFT, see, e.g., Aydemir, Anber, and Donoghue (2012) and Calmet and Casadio (2014).

The approximate truth claim states that an effective theory provides some accurate representations of unobservable physical objects specifiable within the limited range where the theory is likely to remain empirically accurate—or, at least, that we can construct such representations by modifying the original structure of the theory.[16] Again, the argument is relatively standard for the realist: (i) the best explanation for the predictive success of the theory within some regime is that the theory is approximately true; (ii) the probable predictive failure of the theory beyond its limiting scale gives good reasons to take only the descriptions below this scale to be approximately true. In Example 2, for instance, we should expect the descriptions of the dynamical properties of the field to be approximately true if they are restricted to scales lower than Λ. We can also impose limits at large distances by introducing a low-energy cut-off. And one way to construct a model satisfying this restricted set of descriptions is to replace the standard Minkowski space-time with a space-time lattice of finite extent (a sharp low-energy cut-off) and non-zero spacing (a sharp high-energy cut-off) and represent the quantum field in terms of a lattice field defined by assigning a variable $\phi(x)$ to each point of the space-time lattice. As we will see in section 3.5, the approximate truth claim does not mean that, in its standard formulation, an effective theory always accurately represents the putative objects specified by a literal interpretation of its core descriptions. And in section 3.6, we will see that the approximate truth claim sits in tension with other realist requirements in the context of QFT.

4. The stability claim: The representations of an effective theory specified within its domain of empirical validity are likely to remain approximately accurate under theory-change (cf., e.g., Cao and Schweber, 1993, sec. 4.1, sec. 4.3; Wallace, 2006, sec. 3.2-.3.3; J. D. Fraser 2018, sec. 3-4; Williams, 2019b, sec. 3).

Here the challenge is that a future higher-level or same-level theory might undermine the putative representational achievements of our best effective theories. As we will briefly see in section

[16] I will set aside issues related to the nature of scientific representation and use interchangeably “approximately accurate representation" and “approximately true description", assuming that a description is approximately true relative to the actual world if it is satisfied by some model that provides an approximately accurate representation of some actual target system.

3.6, Williams (2019b) and J. D. Fraser (2018; 2020b) rely on the machinery of EFTs, including Wilsonian Renormalization Group (RG) methods, to defend the stability claim in the context of QFT. If we move outside of this context, we can still gain some support for this claim by focusing on the role of higher-order terms in effective theories.

Consider Example 1 and suppose that the predictions of the effective Newtonian theory with a few lowest-order terms are accurate at large distances $r \gg r_0$. If we discover a radically new theory revealing that the predictions of the effective theory are slightly inaccurate at large distances, we can always add higher-order terms to compensate for these empirical discrepancies. This move is, of course, largely *ad hoc*. But it shows that the higher-order terms can be used to encode the contributions of new physics at large distances according to their relevance and thus suggests that these terms do not simply correspond to arbitrary modifications of the theory, with no physical significance whatsoever. The ability of higher-order terms to stand for fine-grained features of new physics is also supported by explicit derivations of effective theories from more comprehensive ones (see, e.g., Eq. 3.6 above). And, in general, the structure of an effective theory is such that we can parametrize the contributions of any type of new physics at large distances up to an arbitrarily high degree of precision by adding increasingly many terms depending only on the degrees of freedom of the original theory. In the Newtonian case, we can even include such terms by preserving *all* the core principles of the original theory (e.g., the structure of the classical Newtonian background space-time and Galilean invariance).

Now, the crucial point is that the contributions of the higher-order terms become increasingly negligible at large distances $r \gg r_0$, no matter what the new physics looks like. And insofar as these higher-order terms are assumed to stand for fine-grained features of new physics, this shows that the descriptions of the effective theory which are relevant at large distances are largely insensitive to the particular details of the new physics. This new physics affects at most the value of the parameters of the lowest-order terms. At the scale r_0, by contrast, the core principles of the effective theory do not even allow us to give an approximately true description of the dynamical behavior of the system and we have no choice but to look for a new theory.

Of course, the previous argument is far from fully ensuring that the theoretical content of some effective theory will not be found to be radically incompatible with the theoretical content of some future theory, even within its domain of empirical validity (see Ruetsche, 2018; J. D. Fraser 2020b, pp. 13-14, for a similar worry). One might also raise legitimate doubts about the ability of the higher-order terms to adequately encode the entirety of the new physics relevant at large distances. Giving a full response to these worries goes beyond the scope of this chapter. If we leave them aside, the previous argument still goes some way toward giving us confidence in the robustness of the theoretical content of the effective theory within its domain of empirical validity.

To summarize, the Standard Effective Account takes effective theories to make approximately true and stable claims about a limited part of the world beyond which it is reasonable to expect to discover (or beyond which we have already discovered) new entities or structures. Although more work needs to be done in order to give a full defense of these features, they suggest nonetheless that effective theories provide us with a reliable epistemic standpoint to identify unobservable entities or structures in the regimes where our best theories are known to be successful. This extends Williams and J. D. Fraser's recent claim beyond the context of QFT and provides a further response to philosophers who deem EFTs unfit for interpretative purposes (e.g., D. Fraser, 2009; 2011; Kuhlmann, 2010). And if we are to interpret effective theories in realist terms, their structure provides us with one central constraint for making more reliable ontological commitments than those commonly made across physics: namely, we should only commit to the existence of concrete physical objects—entities, structures, properties, quantities, states, phenomena, dispositions, and so on—specifiable within the domain of empirical validity of the theory. Beyond this domain, the structure of effective theories gives us good reasons to believe that they fail to represent, or misrepresent, physical objects.

3.5 A Challenge for the Traditional Realist

I will now illustrate how effective theories force the traditional scientific realist to be more selective about her ontological commitments than she might think she has good reasons to be.

Suppose for the sake of the argument that our realist feels unmoved by the traditional constructive empiricist concerns about unobservables and underdetermination (van Fraassen, 1980), the traditional pessimistic meta-induction argument (Laudan, 1981), and the more recent problem of unconceived alternatives (Stanford, 2006). Yet, impressed by the new dogma of effective theories, our realist concedes that our best current theories are best understood and formulated as effective theories and agrees to endorse the account developed in sections 3.3-3.4. She examines the standard formulation of our best effective theories (e.g., the Standard Model Effective Field Theory), either eliminates or disregards their artifactual mathematical structures (e.g., gauge redundancies), and, after interpreting the remaining core theoretical descriptions in their literal sense as she has always done, finds out that our best effective theories represent putative infinite entities and structures, including continuum quantum fields and their infinitary symmetry structure. Of course, our best effective theories might be superseded one day, perhaps by some advanced type of effective string theory or maybe even by some final theory, and our realist is ready to grant that these putative entities and structures are only approximately similar to more fundamental ones. I will restrict myself to entities for simplicity and argue that, on the Standard Effective Account, our realist is not even warranted in taking the representations of these putative entities to be approximately accurate and cannot, therefore, reliably commit to their existence.

Two important clarifications are in order. First, I take an infinite representation to be any type of infinitary model which stands for a putative infinite physical system, i.e., a system with at least one core property specified by a constant or a parameter which takes an infinite value (e.g., a system with an infinite number of degrees of freedom or a wire of infinite extent). Second, I will assume that standard mathematical means of comparison (e.g., measure, cardinality, isomorphisms, etc.) provide reliable standards of relative similarity and accuracy as it is usually assumed in the literature (e.g., da Costa and French, 2003; Weisberg, 2013, chap. 8). I will also first rely on a general notion of similarity in the argument below and then use the specific case of the model-theoretic account of similarity to make the argument more concrete.

How should we evaluate infinite representations then? First, note that there is no specific issue if we believe that our best current theories should be interpreted as offering putatively complete and universal descriptions of the world (e.g., Earman and Roberts, 1999, pp. 445-6).[17] For if these theories are consistent and defeat appropriate competitors, and if the infinite representation of interest is physically motivated, anchored into the mathematical structure of the theory, and irreducible to a mere mathematical artifact, we might take the success of the theory to be a good enough reason to commit to the existence of the corresponding infinite physical system. The Standard Effective Account suggests a radically different conclusion: an infinite representation always represents core features of a putative system in domains where the corresponding effective theory provides, as a matter of principle, unreliable information. I briefly justified the claim that infinite representations are strictly inaccurate—and hence best understood as infinite idealizations—with the help of Example 2 in section 3.4. Continuing with this example, let me now explain why infinite representations are not even close to being approximately accurate.

Recall from the approximate truth claim that, for each effective theory, we can at least construct one realistic finite representation R_f of its target system (i.e., accurate and specified within the domain of empirical validity of the theory). In the Newtonian case, for instance, we can represent the target system in terms of a set of sufficiently large massive bodies moving at non-relativistic velocities with three degrees of freedom to track the center of mass of each body and with a set of finite constants and variable parameters with limited range to characterize their properties (e.g., size, mass). In the ϕ^4-theory case, we can represent the target system in terms of a lattice field defined by assigning one degree of freedom to each point of a space-time lattice of finite size and non-zero spacing. Of course, in the same way as we do not need to reduce a massive body to its point-like center of mass, we do not need to assume that the target system in the ϕ^4-theory takes the form of a “grid". A representation is approximately accurate if the putative entities specified by the representation are approximately similar to real ones. A representation which only ignores,

[17] A similar attitude is expressed, for instance, by D. Fraser when she claims that “there is a crucial difference between QSM [Quantum Statistical Mechanics] and QFT with an infinite number of degrees of freedom [...]: whereas the description of a system as containing an infinite number of particles furnished by QSM is taken to be false, the description of space as continuous and infinite that is furnished by QFT with an infinite number of degrees of freedom is taken to be true" (2009, p. 565).

omits, or abstracts away irrelevant features of the target system does not necessarily provide false information—the only thing we can be certain of is that it provides partial information about the target system.

Now, suppose that for the effective theory of interest, we are also able to construct an infinite representation R_∞ closely related to R_f. For instance, in Example 2, we can decrease the lattice spacing, increase the size of the lattice, and attribute a new degree of freedom to every newly added space-time point in the set specifying the elementary structure of the lattice. However, the more we replace, add, or distort features of the target system in sufficiently small regions of space-time, i.e., the more we take into account descriptions assigning properties to the target system beyond the limits of empirical validity of the effective theory, the more the theory provides false information about the target system. In the limit, the lattice field is replaced by a continuum field with an infinite number of degrees of freedom, and the infinite representation provides us with an infinite amount of false information about the target system in arbitrarily small regions of space-time compared to R_f. The Standard Effective Account thus gives us principled reasons to believe that the infinite representation fails to be even approximately accurate.

We can make the argument more concrete by relying on a specific notion of similarity.[18] According to the model-theoretic (or structuralist) account, for instance, two representations, or mathematical structures in this case, are similar to one another if they are isomorphic to one another, i.e., roughly speaking, if the two mathematical structures have the same number of elements and the same structural relations between their elements. Obviously, an infinite representation is not isomorphic to a finite representation; but few philosophers actually think that the traditional notion of isomorphism provides an adequate standard of accuracy and the problem is to define an adequate notion of “approximate isomorphism". da Costa and French (2003) suggest the notion of

[18] I doubt that the argument actually depends on one's favored account of similarity if we assess whether an infinite representation itself (and not some finite representation thereof) is similar to another finite representation. In the contrast-account, for instance, we need to evaluate the amount of properties shared by two representations and subtract the properties that differ between them, with specific weights assigned depending on whether the property is deemed more or less relevant (see, e.g., Weisberg, 2013, chap. 8, for a recent defense of this account). If we want to compare different fields themselves (and not simply their configurations), the number and type of their degrees of freedom appear to be essential, which means that, according to the contrast-account, two lattice fields with different spacing will be again much more similar to one another than either will be to the corresponding continuum field (cf. below).

"partial isomorphism" (or "partial homomorphism"): briefly put, two mathematical structures M_1 and M_2 are partially isomorphic to one another if there is some mapping from the elements of M_1 to the elements of M_2 which preserves the substructures (and absence thereof) holding between the elements in M_1 and which does not say anything specific if we do not know whether some substructure holds or not between the elements in M_1 (see, e.g., da Costa and French, 2003; Bueno and French, 2011, for more details).

Clearly, it is essential that the two models have important chunks of substructures in common for them to be approximately similar to one another. In this case, two finite representations will always be much more partially isomorphic (or homomorphic) to one another than either of them will be to the corresponding infinite representation. It is non-trivial to give a precise account of degrees of partial isomorphism (or homomorphism) and I will restrict myself to giving an intuitive picture. In Fig. 3.1, for instance, the two lattice fields at the top have, respectively, 64 and 49 elements and share a large part of their spatial structure. We could also specify the substructures which are not preserved (e.g., the local rotational symmetry transformations of the elements which leave the lattice invariant) and the substructures for which we do not know whether they are preserved (e.g., some relations not depicted in the pictures). In contrast, the continuum field depicted in the top right-hand corner has infinitely many more elements than the two lattice fields and infinitely many spatial relations not reflected in the spatial structure of the two lattice fields. Agreed, the patterns of the continuum field might represent well some patterns of the lattice fields (see Fig. 3.1, bottom). But this does not affect the conclusion that the two lattice fields themselves are much more similar to one another than either of them is to the continuum field. We should not underestimate the difference of size and structure between finite and infinite systems.

Let me conclude this section with two comments before examining Williams's and J. D. Fraser's defense of selective realism in QFT. First, the argument above applies to the standard formulation of our best effective theories, and therefore offers a concrete challenge to the traditional scientific realist insofar as he is willing to make ontological commitments by interpreting the central parts of our most successful theories in their literal sense. Second, the argument crucially relies on the

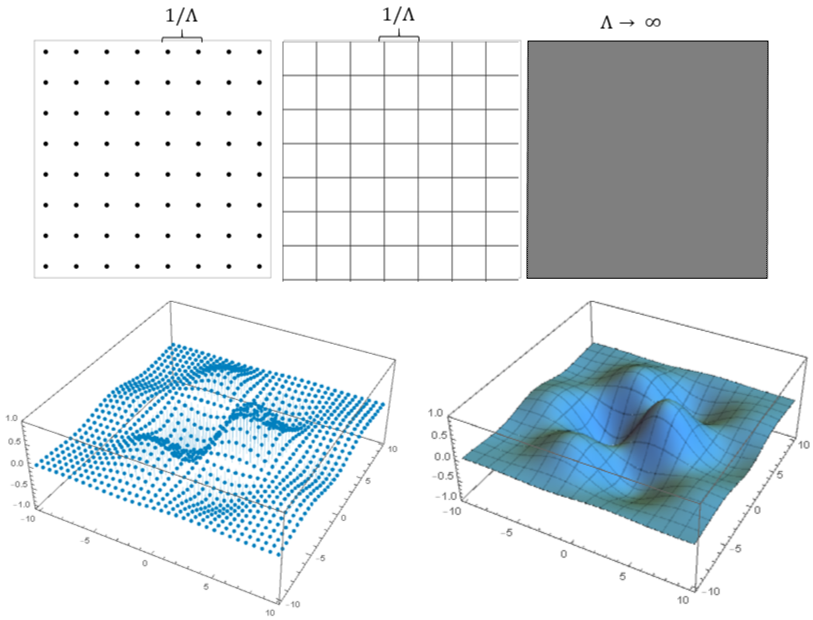

Figure 3.1: Schematic representations of a lattice field and a continuum field, with Λ a sharp cut-off. The two figures in the top left-hand corner represent, respectively, a finite set of points separated by a characteristic distance Λ and a finite set of blocks of characteristic size Λ. The figure in the top right-hand corner represents a continuum of points. The bottom figures represent, respectively, a lattice field configuration and its continuum counterpart.

structure of effective theories. If we have external reasons to believe that our best theories at a given time are likely to be empirically inaccurate at some scale, we might still believe that these theories give approximately true descriptions of more fundamental entities and structures. For instance, we might believe that a low-energy continuum field theory provides an approximately accurate representation of the continuum field described by a more fundamental high-energy theory. The structure of effective theories prevents us from holding such beliefs, no matter what the new high-energy physics looks like.

3.6 Effective Field Theories and Selective Realism

We have seen that effective theories force us to adopt a differentiated attitude towards the entities and structures that we can reliably admit in the realist inventory. In particular, we cannot admit entities if their core properties are specified in regimes where the predictions of the effective theory of interest are likely to break down. Yet, these restrictions leave, in principle, ample space for making reliable and distinctively realist ontological commitments. In the Newtonian case, for instance, we can commit to the existence of sufficiently large massive bodies of center of mass $x_i(t)$ orbiting at sufficiently large distances from each another and moving at sufficiently low velocities, including black holes which, I take it, qualify as unobservables according to van Fraassen's original distinction (e.g., 1980, pp. 13-9). I will now argue that, in some specific theoretical contexts including classical and quantum field theory, the restrictions imposed by the structure of effective theories still raise a challenge for more refined forms of scientific realism. To bring the point home, I will focus on Williams's (2019b) and J. D. Fraser's (2018; 2020b) defense of selective realism in the context of QFT.[19]

The strategy of the selective realist is to defend the realist cause by conceding that our best theories do not get everything right and isolating their parts which both play an essential role in their explanatory and empirical success and are likely to be preserved under theory-change (see, e.g., Psillos, 1999; Chakravartty, 2007). Upon entering the realm of QFTs, the selective realist counts herself doubly fortunate, at least at first sight. First, she can use EFT methods to formulate and interpret our best current theories in a more epistemically reliable way. She has, in particular, efficient tools for evaluating the contributions of a theory in different regimes and eliminating, or "integrating out", its theoretical constituents which are irrelevant in the regimes she is interested in. Second, she can also use the resources of renormalization theory and, in particular, the Wilsonian RG in order to analyze the scale-dependent structure of our best EFTs and increase her confidence in the robustness of their low-energy theoretical descriptions. It is beyond the scope of this chapter to give a detailed account of Wilsonian RG methods (for a recent review, see Williams, 2019a). Here, I will restrict myself to discussing the interpretative constraints that Williams and J. D. Fraser extract from EFT and RG methods and evaluating the success of their selective strategy.[20]

How, then, should we separate the theoretical descriptions of our best current EFTs if we want

[19] I will leave aside Wallace's account insofar as he is primarily concerned with defending the foundational and interpretative relevance of cut-off Lagrangian QFTs in (2006; 2011) and not scientific realism strictly speaking (or, more precisely, structural realism).

[20] See also Ruetsche (2018, 2020), Rosaler and Harlander (2019, sec. 5.6), and Rivat and Grinbaum (2020).

to implement the selective realist strategy? Since the structure of an EFT gives us good reasons to believe that its predictions break down at some high-energy scale, we should first restrict our attention to the parts of the theory which describe its low-energy content:

1. Isolate theoretical descriptions which are specified within the limited range of scales where the theory is likely to remain reliable (see, e.g., Williams, 2019b, p. 13).

As already discussed in section 3.4, constraint 1 purely follows from the structure of effective theories.

Some of these low-energy descriptions might still depend significantly on irrelevant parts of the theory or involve representational artifacts (e.g., the specific type of cut-off in Eq. 3.4). We need, therefore, to introduce further constraints if we want to isolate the parts of the theory which play an essential role in its explanatory and predictive success and which accomplish genuine representational work. Williams and J. D. Fraser remain somewhat ambiguous here. They highlight various ways in which EFT and Wilsonian RG methods allow us to gain confidence in the "robustness" of the low-energy content of EFTs. Yet, they also appear to put emphasis on two different robustness criteria. Williams seems to be more concerned with the relative insensitivity of the low-energy physics to the high-energy physics:

> [...] it is one of the essential virtues of the RG that it provides a tool for determining how changes in the structure of the theory at the scale of the short-distance breakdown affect physics at longer distances where the theory is empirically reliable. What the RG shows is that the 'fundamental' short-distance structure with which standard interpreters are so concerned is largely irrelevant to the physical content of an EFT in the domain where we have any reason to consider it empirically reliable. (2019b, p. 16)

J. D. Fraser, by contrast, puts emphasis on a more general type of invariance, which includes the mathematical invariance of the low-energy descriptions of the theory under different parametrizations and other representational artifacts introduced when renormalizing the theory (e.g., J. D.

Fraser, 2020b, p. 12; 2018, p. 1172; see also Ruetsche, 2018, pp. 11-2; 2020, p. 16; Rosaler and Harlander, 2019, sec. 5.6).

Despite important overlaps, as we will see below, I think that it is crucial to distinguish between two main interpretative constraints to account for Williams's and J. D. Fraser's slightly different outlooks and for the variety of ways in which the low-energy content of an EFT amenable to RG methods is robust:

2. Isolate theoretical descriptions which are largely insensitive to high-energy physics;

3. Isolate theoretical descriptions which are invariant under RG-transformations and independent of specific choices of renormalization methods.

Constraint 2 is mainly derived from the structure of effective theories, although RG methods often allow us to refine the analysis. As we saw above, part of what makes an effective theory distinctive is that its descriptions which are significant within a specific regime are largely independent of its descriptions which are significant within a different regime (e.g., lower- vs. higher-order interaction terms in Examples 1-2; light vs. heavy field dynamics in Eq. 3.6). In particular, it is usually possible to modify the high-energy content of an EFT without affecting much its low-energy content, including its low-energy predictions (e.g., by adding higher-order interaction terms in Examples 1-2). We can also usually show that different high-energy theories reduce to the same low-energy theory, or at least to similar ones (e.g., we can add a third heavy scalar field in Eq. 3.6 and obtain a similar low-energy theory after integrating out the two heavy fields and making appropriate approximations). In all these cases, the crucial point is that the low-energy content of the theory is robust under variations of its high-energy content. And, in general, the bulk of the low-energy content of the effective theory depends only on a finite number of free parameters (see Examples 1-2).

Constraint 3, by contrast, arises specifically from a RG analysis. In general, a theory can be renormalized in many different ways, and the specific renormalization method chosen usually requires us to introduce some arbitrary scale parameter (e.g., the parameter Λ in Example 2) and use some particular scheme to absorb the terms depending on this parameter (e.g., a mass-dependent renormalization scheme). Thus, constraint 3 requires us to isolate theoretical descriptions which are invariant under different renormalization methods and choices of scales (cf., Williams, 2019b, p. 12; J. D. Fraser, 2018, p. 1172; 2020b, p. 12).

We can, in fact, look at this constraint in two distinct ways. (i) If we consider some fixed high-energy theory, we can derive a series of low-energy theories by successively integrating out high-energy field configurations in their path integral formulation. In this case, constraint 3 is best understood as requiring us to isolate invariant theoretical descriptions in the series of low-energy theories. (ii) If we consider some low-energy theory with parameters fixed by means of experimental inputs, we can show that this theory and its parameters remain unaffected by changes in the high-energy theory from which it is originally derived, i.e., the so-called "bare" theory (cf. Wallace, 2006, p. 49; 2011, p. 6; Williams, 2019b, p. 12; J. D. Fraser, 2018, p. 1172; 2020b, p. 12). In this case, constraint 3 is best understood as requiring us to isolate theoretical descriptions which are not affected by changes in the value of the high-energy cut-off and in the parametrization of the high-energy theory.

Now, in addition to adopting constraints 1-3, the selective realist also needs to make sure that she is offering a genuine defense of the realist cause. First, in order to give a sufficiently informative and non-ambiguous explanation of the success of the theory, she needs to isolate a definite set of unobservable entities or structures with clear identity conditions—say, in the case of entities, with a well-specified set of core properties which distinguish them from other entities, whether they are fundamental or not. For instance, in the Newtonian case, we might identify a system by means of its position, its velocity, its mass, and its dynamical behavior. If we simply give a functional characterization of the system by means of its mass, for instance, we are likely to pick out very different types of entities and leave the target of our commitments indeterminate. Likewise, in the QFT case, we might identify a system by means of the type and number of its degrees of freedom, its mass, its self-interacting parameters, and its dynamical behavior. If we simply specify the system by means of its dynamical behavior and its mass, for instance, there is still some ambiguity as to whether we pick out a lattice or a continuum field. Contrary to what Williams (2019b, p. 15) suggests, to simply "extract reliable ontological information" does not suffice (see also J. D. Fraser, 2020b, p. 12). The selective realist needs to give a sufficiently comprehensive account of a definite set of entities or structures in order to fulfill her explanatory duties.

Second, the selective realist needs to give a literal interpretation of some privileged parts of the theory, as it is often assumed in the literature (e.g., Psillos, 1999; Chakravartty, 2007). In the Newtonian case, for instance, the selective realist can take the theory to literally describe a black hole with a center of mass specified by the position $x(t)$ and which interacts gravitationally with other bodies. The gravitational force can be interpreted as a concrete structure, i.e., as a variable relation with a specific strength depending on the relative position and the masses of the bodies. Although Williams and J. D. Fraser do not give much details about their preferred version of selective realism, they both seem to endorse this semantic constraint, i.e., that the privileged set of descriptions that we take be trustworthy should be understood in their standard sense and taken to be approximately true or false *simpliciter*.[21] In the same vein, the selective realist should avoid modifying too much the original mathematical structure of the theory or engaging into any other form of *post hoc* interpretative practice. Otherwise, she will fail to take the original theory at face value and explain its explanatory and predictive success in its own terms. This is well illustrated, for instance, by attempts to draw conclusions about the ontological content of our best current QFTs based on their putative algebraic reformulation, despite the fact that they have not yet been successfully formulated in algebraic terms.[22]

The difficulty now is that it is not clear what the selective realist should commit to if she endorses these constraints in the case of our best current EFTs, as it has been acknowledged by J. D. Fraser (2018, p. 1172; 2020b, p. 15). I will expand on Ruetsche's recent discussion in

[21] At the very least, this seems to be implicit in the central question underlying Williams's and J. D. Fraser's interpretative stance—"given that this theory provides an approximately true description of our world, what is our world approximately like?" (Williams, 2019b, p. 2). Reference to particular physical scales seems to be included in the properties of the target system (see, e.g., J. D. Fraser's reference to the "bulk properties" of a fluid when he illustrates the idea of large distance features of the world, 2018, p. 1173).

[22] See, e.g., D. Fraser (2008) for such an attempt and Williams (2019b) for a criticism, emphasizing the importance of paying attention to how QFTs are successfully implemented in practice.

(2018; 2020) by looking at the most obvious candidates—correlations, particles, and lattices—and argue that they do not allow us to meet constraints 1-3 or make distinctively realist ontological commitments.

Correlations: J. D. Fraser proposes to focus on low-energy correlation functions:

> [...] a preliminary strategy is to point to correlation functions over distances much longer than the cutoff scale as appropriate targets for realist commitment. These quantities are preserved by the renormalization group coarse-graining transformation and encode the long distance structure of a QFT model. They are also directly connected to its successful predictions—you cannot vary the long distance correlation functions of a theory without drastically affecting its low energy scattering cross sections. (2018, p. 1172)

We face several issues here. First, it is not clear how we should interpret correlation functions. In the standard QFT framework, they correspond to vacuum expectation values of time-ordered products of field operators at different space-time points. The simplest textbook interpretation in the simple case of two field operators $\hat{\phi}(x)$ and $\hat{\phi}(y)$ is to take the expectation value $\langle 0|\mathcal{T}\{\hat{\phi}(x)\hat{\phi}(y)\}|0\rangle$ to measure the probability (once squared) that a particle is created at some earlier point x, propagates, and is annihilated at some later point y (assuming $x_0 < y_0$). This interpretation is controversial, in large part because of the difficulties associated with the interpretation of quantum fields and particles in interacting QFTs. The crucial point here is that however we interpret these entities (I discuss the two cases below), we need to commit to something more than correlations if we follow this standard textbook interpretation. Likewise, if we interpret correlation functions more generally as standing for the degrees of co-variation or coordination between two variables at two distinct points, we need to commit to something more than degrees of co-variation (I discuss the case of physical degrees of freedom below).

We might opt for a more minimal interpretation of correlation functions as encoding structural physical information independently of the physical objects or variables they relate. In the case of EFTs, we can interpret correlation functions as encoding the correlations of the target system at sufficiently large distances, where 'correlation' refers to a set of numbers characterizing the degree of correlation between two space-time points or regions. If we take this path, however, the empiricist might raise doubts about the distinctively realist character of these commitments and, instead of rejecting them altogether as she usually does, simply re-appropriate them as her own as Ruetsche (2020, pp. 16-7) rightly notes. It turns out that the framework of QFT even gives her good reasons to do so. Typically, in high energy physics, we summarize empirical information about the correlations between the initial and final states of some scattering process in a mathematical object called the S-matrix, and the S-matrix can be derived by taking the appropriate asymptotic limit of a sum over all the possible correlations between initial and final states by means of the LSZ reduction formula (see, e.g., Schwartz, 2013, sec. 6.1). If we take the state of a field to be in principle observable in any sufficiently large region of space-time, nothing seems to prevent the empiricist from understanding the numbers specified by correlation functions as simply summarizing the empirical information that *would* be gathered about the correlations between two states of the system if we were to make measurements in this space-time region.

Even if the structural realist finds a way of avoiding this empiricist re-appropriation, she still faces one important issue. Strictly speaking, correlation functions in QFT are not RG-invariant contrary to what J. D. Fraser claims. If we implement a coarse-graining procedure by integrating out high-energy field configurations, for instance, the different correlation functions obtained at low energies are multiplied by "wave function normalization" factors. In general, these multiplicative factors depend on other variables, such as the couplings of the theory. And so it does not appear that there is an invariant and therefore unambiguous characterization of the degree of correlation between two distinct space-time points since it depends on the way we parametrize the low-energy theory. By contrast, S-matrix elements are invariant under these different parametrizations. Similarly, the path integral used to generate the set of correlation functions is also invariant under different coarse-graining procedures. Yet, it seems to be even more difficult to interpret the S-matrix and the path integral in distinctively realist terms compared to correlation functions. And, again, the empiricist might simply re-interpret the S-matrix and the path integral as bookkeeping devices for all the possible empirical information that we could gather about the correlations between initial and final states of the system in sufficiently large space-time regions.

Particles: Another option, perhaps more likely to enable us to make distinctively realist ontological commitments, is to focus on particles, such as protons, neutrons, gluons, and photons (see, e.g., Williams, 2019b, p. 20, p. 22). The concept of particle in interacting QFTs which involve an infinite number of degrees of freedom is controversial (see, e.g., Teller, 1995; Bain, 2000; D. Fraser, 2008; Ruetsche, 2011). In the modern understanding of QFT, it is common to understand particles in terms of patterns of excitations in the fields (as it is rightly noted by Wallace, 2006; 2019, sec. 4, for instance). This understanding is robust whether we deal with the perturbative or exact, non-interacting or interacting formulation of a QFT with an infinite or finite number of degrees of freedom (ignoring the mathematical issues inherent in realistic continuum QFTs). And, to be more precise, we can interpret particles in terms of sufficiently well-behaved and localized patterns in the field configurations in regimes where the interactions described by the theory are sufficiently weak.

Again, the main problem here is that neither field configurations nor energy-momentum states are RG-invariant. In general, RG-transformations mix both field operators and the states of different kinds of particles with one another. The only notion of "particle" that does not suffer from these issues is the one specified by the asymptotic states in the non-interacting version of the theory. But insofar as we seek a realist interpretation of *interacting* QFTs, we cannot simply restrict our commitments to the free particles that we prepare and detect in experiments. And even if we were to take this extreme route and leave aside potential empiricist re-appropriations, we would still not be able to commit to the existence of particles such as quarks and gluons insofar as the quark and gluon fields do not have asymptotic elementary particle states.

Lattice fields: A third option is to focus on low-energy degrees of freedom (e.g., as represented by the field operators associated with the variables $\tilde{\phi}(k)$ for $k \ll \Lambda$ in Example 2). Agreed, many of the properties associated with these degrees of freedom do vary under RG-transformations, including coupling parameters and the specific form of the variables used to specify these degrees of freedom (which depends, in particular, on how we separate low- and high-energy degrees of freedom). Yet, whether we integrate out a large or a small range of high-energy field configurations, the number of degrees of freedom at sufficiently low energies remains exactly invariant. We could, therefore, consider them to be an appropriate target for the selective realist, as Williams sometimes seems to suggest (2019b, p. 13, pp. 14-5). The main problem here is that this might not be enough for the realist. We can interpret a degree of freedom as a determinable dynamical property of some system. However, without a specification of the low-energy system, any appeal to low-energy degrees of freedom will remain too indeterminate for the realist and therefore undermine her attempt to provide a sufficiently informative and unambiguous explanation of the success of the theory. After all, these degrees of freedom could perfectly stand for the properties of radically different low-energy systems. They could be, for instance, the degrees of freedom of low-energy lattice fields with different types of spatial structures.

In order to avoid the issue of underdetermination at low energies, we can perhaps isolate a privileged set of low-energy lattice fields for our best current EFTs. If we put a given EFT on a lattice of finite size and spacing, we can indeed integrate out high-energy degrees of freedom, obtain low-energy lattices, and eventually derive empirically equivalent low-energy *predictions* which do not significantly depend on the details of the short-distance physics and on the way we eliminate high-energy degrees of freedom (cf. Wallace, 2006, pp. 48-50). In addition, these low-energy lattices are well-specified within the limited range of energy scales where the EFT of interest is likely to remain reliable, and they do appear to enable us to make distinctively realist ontological commitments.

Yet, we still face a severe issue of underdetermination both at low and high energies. If we formulate an EFT on a lattice and interpret its low-energy descriptions in their literal sense, the RG coarse-graining transformations appear to force us to commit to the existence of different lattice fields at different low-energy scales. We might solve this issue by claiming that these lattice fields are more or less coarse-grained partial instantiations of the same high-energy lattice field. If we fix any of the low-energy lattice representations, however, RG methods allow us to change the high-energy lattice representation without affecting the low-energy lattice one. And this introduces some pernicious form of underdetermination about what the low-energy lattice representations are supposed to stand for.

There are two additional points that make the matter even worse. First, if we start with a given lattice field, we can implement a specific type of coarse-graining procedure that defines a lattice field with a different number of degrees of freedom but with the same lattice spacing. We simply need to rescale the original lattice spacing and adjust the parameters of the theory after having integrated out high-energy degrees of freedom. And the two lattice field representations are, of course, empirically equivalent (see, e.g., Hollowood, 2013, sec. 1.2, for a simple explanation of this specific way of implementing RG-transformations). Second, the specific form of the low-energy lattice representations depends on the type of coarse-graining procedure we implement in the first place. We might separate low- and high-energy degrees of freedom in very different ways, or define new low-energy degrees of freedom by averaging over high-energy ones in a particular way. In each case, the procedure yields a different set of low-energy lattices. And overall, then, it appears that low-energy lattices do not allow us to satisfy constraint 3.

Now, if we are to make distinctively realist ontological commitments about entities or structures in the case of our best current EFTs and maintain Williams's and J. D. Fraser' robustness constraints, continuum quantum fields appear to be ideal candidates. Assuming that we do not latticize the theory, we may either take a smooth cut-off or a sharp cut-off (in which case we eliminate high-energy states of the field), and keep higher-order interaction terms or eliminate them (depending on the desired accuracy). Either way, the theory describes a RG-invariant continuous system with an infinite number of degrees of freedom, at least for a finite range of scales. If we keep all the degrees of freedom in the theory, we do not face the issues encountered with lattices. And if we do not focus on the specific values of the properties of the continuum field, such as the value of its mass, the strength of its interactions, or the value of its field configurations on space-time, we also avoid the issues encountered with correlation functions and particles.

The main issue here comes from the domain restriction claim.[23] On the face of it, we are committing to entities with core properties specified in regimes where the predictions of the EFT of interest are likely to break down, and this should be a good enough reason not to make such commitments (as Williams and J. D. Fraser would probably agree). In response, we might insist that we are committing to the existence of continuum quantum fields insofar as they are approximately similar to large distance scale features of the world. If we wish to endorse the literalness constraint, however, we cannot make such a claim. As we saw in section 3.5, if we take the descriptions of a continuum quantum field itself at face value, i.e., as being either (approximately) true or false, we are forced to attribute degrees of freedom to some putative entity in arbitrarily small regions of space-time, and the structure of effective theories gives us reasonable grounds not to commit to the existence of such entities.

We might also try to escape the difficulty by taking the infinite representation of the putative continuum field to contain a finite part that does the appropriate representational work at large distances, say, a finite representation of a lattice field. The problem here is that any specification of such finite representation involves a particular specification of an arbitrary lattice spacing, or at least of a finite number of degrees of freedom, and therefore brings us back to the issues discussed above. The best RG-invariant representations of putative entities in our best current EFTs appear to be the representations of continuous systems with an infinite number of degrees of freedom. And we cannot simply embed these representations in finite ones without losing their representational value altogether.

3.7 Conclusion

I will briefly conclude with a more radical suggestion to defend the realist cause in the case of our best current EFTs. To summarize the main points of the chapter first, we have seen that the structure of effective theories across physics is best characterized by the fact that they contain

[23] Another set of issues that I will not discuss here is related to the existence of empirically equivalent field representations (for a discussion about Borchers classes, for instance, see Haag, 1996, sec. II.5.5; Wallace, 2006, sec. 2.2, 3.3).

intrinsic empirical limitations. In a slogan: effective theories "predict" their own predictive failure at some scale. We have also seen that the most straightforward realist explanation of this predictive pattern is to take effective theories to accurately represent limited parts of the world, which provides one central constraint for the sort of entities and structures that a realist might reliably include in his inventory if he takes effective theories seriously. I gave one concrete example of the sort of entities that the traditional scientific realist cannot commit to if he interprets the core descriptions of effective theories in literal terms: namely, he cannot commit to the existence of putative infinite systems since their individuating properties are specified in regimes where the predictions of the theory are likely to break down. Yet, the domain of empirical validity of an effective theory leaves, at least in principle, enough space for the realist to commit to the existence of unobservable entities or structures (as we have seen in the Newtonian case). As I have argued in the last section, this is not always straightforward. In particular, the structure of our best current EFTs is such that it is not clear what we should commit to if we want to make distinctively realist ontological commitments and avoid making these commitments depend on irrelevant or artifactual features.

I suspect that many of us still entertain the hope of a robust form of scientific realism that does not totally fail to adhere to its original letter and which is concerned with explaining the success of our best theories in their own terms. In the case of our best current EFTs, a potential candidate for making distinctively realist ontological commitments appears to be continuum quantum fields. And if we want to commit to the existence of such entities at low energies, one potential solution is to modify the traditional semantic tenet of scientific realism (but keep its ontological and epistemological tenets as summarized in, e.g., Psillos, 1999, p. xvii). Instead of taking the descriptions of a continuum field at face value, that is, as being either (approximately) true or false, we need to take them to be (approximately) true or false relative to a specific range of physical scales. That is, when we speak about a continuum field with properties assigned at every point of space-time, we are not literally making the claim that the field has properties at arbitrarily short distances *simpliciter*. We are making a claim about the structure of matter *at* large distances. And the descriptions of an effective theory are approximately true or false relative to these scales up until we discover that the theory breaks down at some limiting scale, in which case we need to work with a new theory. If the new theory is effective, we will be again making claims relative to a specific range of physical scales. This strategy requires us to modify one of the central tenets of scientific realism usually endorsed by selective realists. But it might enable us to explain the success of our best theories in their own terms.

Chapter 4: How Theoretical Terms Effectively Refer

This chapter proposes a new theory of reference to address the problem of referential failure across theory-change. Drawing on Kitcher's and Psillos's accounts, I argue that referential success is best assessed before theory-change by examining whether the central theoretical terms of a theory refer to entities specified within the limited physical context where the theory is empirically reliable. I show that effective theories provide a paradigmatic set-up for implementing this principle of selective reference and serve as a blueprint for assessing the referential success of the usual suspects, such as 'phlogiston' and 'luminiferous ether'.

4.1 Introduction

Many scientific realists share the intuition that our best current scientific theories accurately represent entities that we cannot directly observe in experiments. In support of this intuition, it is often argued that the predictive success of these theories would be hard to explain, or "miraculous" as Putnam (1975) felicitously put it, if they had nothing to do with what the world is like (see also Maxwell, 1962, and Smart, 1963, for early versions of this argument). Yet, this intuition is challenged on many counts and most powerfully by the apparent twists and turns of the history of science. Many of our "best" past theories appear by the light of their successors to be radically false and to contain central theoretical terms which fail to refer to anything real. And this historical pattern, as Laudan (1981) famously argued, does not merely undermine the explanatory link between predictive success, approximate truth and reference dear to scientific realists. It also gives reasons to believe that our best current theories might prove one day to be as radically false and referentially unsuccessful as their predecessors.

The most popular response to this challenge is to concede that our best past theories did not get everything right but maintain that they still contained central parts that survived and thus remain worthy of realist commitments (e.g., Worrall, 1989; Kitcher, 1993; Psillos, 1999; Chakravartty, 2007).[1] Among the “selective" realists who follow most closely the traditional form of scientific realism, such as Kitcher and Psillos, it is often granted that the problem of referential failure across theory-change requires adjusting both one's semantic and epistemic commitments. In particular, if we acknowledge that at least some of the central terms of our best past theories fail to refer to anything real, we cannot simply assume that, in general, the terms of successful theories automatically refer to the right sorts of entities and restrict ourselves to selecting descriptions that we can trust. We also need to account for: (i) the mechanism by which some, but not all, theoretical terms come to refer to unobservable entities; and (ii) the putative referential stability of some, but not all, theoretical terms under theory-change (or their putative referential continuity if the domains of successive theories overlap).[2]

The central challenge underlying both (i) and (ii) is to find a reliable and principled way of distinguishing between referential success and failure, i.e., a principle of selective reference, and this is far from trivial. For instance, we cannot appeal to the theoretical content of our best current theories to assess current and past referential success since we do not yet know whether they will not appear to be deeply mistaken by the light of future theories. Nor can we rely on scientists' judgments since their descendants might prove them wrong. Nor can we point to the crucial predictive and explanatory role of a term since the next theory might show that, ultimately, this term was not playing such a crucial role. We need, in other words, to find a reliable and principled way

[1] In response to Laudan's historical gambit, scientific realists also emphasize that many of his examples are not sufficiently successful to be even considered candidates for realist commitments (e.g., Devitt, 1984, sec. 9.3; McMullin, 1984, p. 17; Worrall, 1989, p. 113). It is often recognized, however, that while this response might work for past theories such as the crystalline spheres theory, other examples such as the phlogiston theory and 19th century theories of the luminiferous ether appear to have enjoyed a sufficient amount of success to pose a genuine threat to realists (e.g., Psillos, 1999, chap. 5).

[2] Structural realists would probably respond that the amount of discontinuity is much less important once we focus on the structural content of our best past theories (e.g., Worrall, 1989; Ladyman, 1998; Ladyman et al., 2007). Even if this proves to be right, both epistemic and ontic structural realists would still benefit from developing a similar account, i.e., an account that: (i) specifies the mechanism by which some, but not all, mathematical equations or structures come to relate to their target; and (ii) explains why some, but not all, mathematical equations or structures are likely to be referentially or representationally stable under theory-change. For simplicity, I will restrict myself to selective strategies closely associated with the traditional form of scientific realism in this chapter.

of assessing referential success *before* theory-change, and adjust the semantics of theories accordingly. I will refer to this challenge as "Stanford's challenge" following Stanford's criticism of the selective realist strategy (2003a,b, 2006, 2015).[3]

The goal of this chapter is to design a theory of reference which allows selective realists to address both the traditional problem of referential failure and Stanford's challenge without making referential success too easy or too hard. I will first engage with Kitcher's (1978, 1993) and Psillos's (1999, 2012) accounts, and argue that they fail to address these issues in a satisfactory manner—although not exactly for the reasons usually raised in the literature (cf., McLeish, 2005; Stanford, 2006; Chakravartty, 2007; Ladyman et al., 2007). Concerning Kitcher, I will argue that his attempt to assess referential success based on the practices and intentions of scientists in different contexts of utterance makes referential success and failure overly sensitive to scientists' idiosyncratic attitudes and willingness to speculate about their subject matter. Concerning Psillos, I will argue that his specific use of causal and descriptive elements of reference-fixing introduces some pernicious form of referential indeterminacy.

Drawing on Kitcher's and Psillos's accounts, I will then propose a theory of reference modeled on the new paradigm of effective theories developed by physicists in the 1970-80s.[4] One of the distinctive features of effective theories is that they contain intrinsic empirical limitations, i.e., their structure incorporates a robust specification of the scales at which they are likely to be empirically inaccurate before we probe these scales in experiments. As a realist, a natural explanation of this predictive pattern is to take effective theories to accurately represent a limited part of the world at best. Hence, instead of focusing on the context of utterance as in Kitcher's theory, I will argue that referential success is best assessed by focusing on the limited objective context delineated by the range of scales where the theory is empirically reliable. I will modify Psillos's theory of reference in light of this principle of selective reference and show that the resulting theory has two main

[3]Stanford also emphasizes another version of this challenge according to which there is no principled way of distinguishing between indispensable and idle parts of a theory before theory-change (see also, e.g., Giere, 1988, p. 96).

[4]For a philosophical discussion of effective theories in the original context of Quantum Field Theory (QFT), see, e.g., Cao and Schweber (1993), Hartmann (2001), and Williams (2019b).

advantages:

(i) It is often, if not always, possible to identify before theory-change at least the limited context where a successful theory has been found to be empirically accurate, which lays the ground for responding to Stanford's challenge. The case of effective theories is special: we can estimate the limiting scale at which the theory is likely to become empirically inaccurate and therefore extend its scope to scales which have not yet been probed in experiments within the range delineated by this limiting scale.[5]

(ii) This sort of limited context provides a reliable basis for selecting the descriptions of the theory which are likely to remain trustworthy under theory-change and therefore identifying stable referents. As we will see, the selection process is remarkably robust in the case of effective theories. In other cases, I will show with the examples of the phlogiston and the luminiferous ether that the framework of effective theories still serves as a blueprint for assessing the referential success of the central terms of a theory.

Two caveats before I begin. (i) One might be worried that this strategy does not handle well cases where the new theory is found to be radically incompatible with the current theory even *within* the range where it is empirically reliable. In response, I will suggest with the examples of 'phlogiston' and 'luminiferous ether' that problematic cases of referential failure in the history of science typically arise because the putative referent of the term at stake is specified outside of the limited context where the theory of interest has been put to the test. For anomalous cases (if any), I will argue that if the referent is characterized by descriptions which are restricted to or associated with this sort of limited context, we still have independent grounds to believe that the term successfully refers both before and after the advent of the new theory. (ii) While the notion of objective context is precise in theoretical physics, it becomes increasingly vague and hard to specify as we move towards the special sciences. So apart from the diagnosis of 'phlogiston', my argument will be restricted to physics. I will nonetheless stretch it as far back as the framework of

[5] In general, estimating the value of these scales requires having appropriate empirical inputs in some accessible regime and assuming that the dimensionless constants of the theory are of order one.

effective theories seems to be applicable in the history of physics by relying on Galileo's account of gravitational phenomena with specific contrasts with Descartes and Newton when needed.[6]

The chapter is organized as follows. Section 4.2 discusses Kitcher's and Psillos's theories. Section 4.3 offers a context-dependent theory of reference modeled on the framework of effective theories. Section 4.4 explains how this theory helps selective realists to address the issue of referential stability and Stanford's challenge. Section 4.5 responds to objections.

4.2 Kitcher and Psillos on Reference

Two competing intuitions are usually at play in the debate over the reference of scientific theoretical terms. Imagine for the sake of the argument that upon the publication of the first edition of the *Principia* in 1687, most British and continental natural philosophers immediately came to endorse Newton's revolutionary conception of gravity as a force acting at a distance between massive bodies (which Newton himself was reluctant to endorse). Imagine further that it was clear to these natural philosophers that the post-scholastic conception of gravity as an intrinsic quality of matter and the Cartesian conception of gravity as arising from the action of material vortices were radically mistaken (despite their dominance at the time). What would be written on the front page of the French magazine *La Gazette*?

An advocate of the description theory of reference would probably write:

> *Revolution across the Channel: Newton discovered universal gravitation and demonstrated that Descartes's gravitational vortices do not exist.*

Here the basic idea is that we pick out entities by means of their properties (see Frege, 1892, and Russell, 1905, for early versions of the description theory). The theoretical term 'gravity' successfully refers if there is some unique entity satisfying the core description associated with the

[6] One might wonder whether Galileo's earlier and later accounts of gravity were sufficiently successful to be of any interest to selective realists (see, e.g., Galilei, 1590, 1632, 1638). Since my goal is to make a claim about the relevance of effective theories for the topic of reference and not about the specific theories that we should include in the realist gambit, I will ignore this worry. For more details about the success of Galileo's accounts of gravity, see, e.g., Koyré (1966, part II) and Drake's introduction in (Galilei, 1638, esp. p. ix).

term and fails to refer if there is nothing satisfying this description. For instance, since Descartes's description of gravity turns out to be radically false, it means that there is no such thing as Cartesian gravity, i.e., gravity as Descartes describes it.

One immediate issue for selective realists tempted by the description theory is that it does not give much leeway for changing the core description associated with a term without losing reference altogether. If we require entities to be picked out by means of a comprehensive set of properties, we are likely to find incompatible sets over time and make the history of science more referentially discontinuous than it appears to us. Inversely, if we attempt to avoid this issue by keeping only a minimal set of properties, or even only observable ones, the core description associated with each term is likely to be satisfied by entities with radically incompatible properties. But referential indeterminacy, i.e., the absence of a matter of fact as to which entity a term uniquely picks out, is not really the sign of referential success. Or, at any rate, we will face some pernicious type of underdetermination as to what makes the theory at stake approximately true if we allow for terms to refer to radically different kinds of entities.

Being aware of these issues, an advocate of the causal or causal-historical theory of reference would probably write instead:

> *Revolution across the Channel: Newton showed that the Cartesians were deeply mistaken about the nature of gravity.*

Here the basic idea is that reference is originally fixed by means of some kind of causal contact with the entity for which a new term is introduced and that referential continuity is ensured during the transfers of the term among competent speakers if the speakers use the term in the same way and intend to refer to the entity originally picked out (see, e.g., Kripke, 1972; Putnam, 1975, for early views).[7] Hence, even if a speaker along the chain decides to characterize the referent of a given term by means of radically new properties, she might still intend to attribute these properties to

[7] Strictly speaking, it is not part of Kripke's and Putnam's views that the reference-fixing event necessarily involves some causal contact. However, they both assume some element of indexicality in this initial event, which seems to require some kind of contact with the actual world (see, e.g., Kripke, 1972, pp. 57-9; Putnam, 1973, pp. 202-4). For simplicity, I will restrict myself to the "full causal theory of reference", to use Kroon's expression (1985, p. 144), which involves causal contact both in the reference-fixing event and the reference-borrowing process.

the same entity picked out by the speaker who first introduced the term. Arguably, then, Descartes and Newton were still talking about the same thing despite holding radically incompatible beliefs about gravity because they were both attempting to characterize the same causal agent responsible for the free fall of terrestrial bodies (among other phenomena).

The causal theory escapes the problems of referential failure and discontinuity that beset the description theory. But it does so at the expense of making referential success too easy. If we introduce a term to pick out the causal origin of a given phenomenon, the term will automatically refer to something insofar as there is presumably always some causal agent for any given set of observed phenomena (Stanford and Kitcher, 2000, p. 115, call this the “no failures of reference" problem). In this case, however, referential success does not depend at all on whether the theory is approximately true, which leaves selective realists tempted by the causal theory in an unstable position. For it means that even if our best theories turn out to be radically false, we would still be able to successfully speak about the fundamental entities of the world (if any) without having any knowledge about them.

In the case of scientific theoretical terms, philosophers usually—and rightly in my view—think that both theories of reference have something right about the mechanism of reference-fixing (see, e.g., Enc, 1976; Sterelny, 1983; Lewis, 1984; Kroon, 1985, 1987, for early views). This suggests a causal-descriptive or “hybrid" theory of reference, the gist of which is that referential success originally depends both on some kind of bare causal link established between a speaker and a referent during a baptismal event and on a privileged set of descriptions used by the speaker to identify this referent. The difficulty, however, is to give a precise account of how the mechanisms of reference-fixing via causal contact and via satisfaction of a description are to be combined with one another. For it seems that we pick out an entity either by pointing at it or by uttering something true—but not by doing both or, at any rate, that the mechanisms of reference-fixing work on their own when we do both. I will call the problem of combining these two mechanisms the “combination problem". It is, of course, a problem for selective realists insofar as they wish to avoid making referential success too hard or too easy, and therefore not a problem for the description and causal theories *per se*. Kitcher's and Psillos's solutions to the problem of referential failure can be seen as two distinct ways of engaging with the combination problem in the first place.

Although Kitcher does not hold exactly the same view over time (see, e.g., Kitcher, 1978, 1982, 1993, 2001; Stanford and Kitcher, 2000), his solution to the combination problem and more generally to the problem of referential failure is to make reference-fixing depend on the context of utterance. For my purpose, the most important claims of his 1978-1993 account can be reconstructed as follows:

(1) Scientists produce different term-tokens of the same term-type, i.e., different instances of the same term, with different reference-fixing mechanisms. Depending on the context of utterance, reference is originally fixed by baptism (causal link) or by description (satisfaction link), and each term-type is associated with a "reference potential", i.e., with the set of ways the reference of the term-type is fixed on different occasions. The referent of a term-type is determined on a token-by-token basis by finding out the "mode of reference" associated with each token.

(2) It is a matter of fact that scientists use the same term to refer to different things on different occasions, i.e., tokens of the same term-type refer non-uniformly once we look closely enough at scientific practice.

(3) Historians of science attribute referential success and failure by coming up with the best explanation for the production of a token on the basis of the "principle of humanity", i.e., by trying to "impute to the speaker whom we are trying to translate a "pattern of relations among beliefs, desires and the world [which is] as similar to ours as possible"" (Kitcher, 1978, p. 534; see Grandy, 1973, for the original reference).

(4) Even if most of the tokens of a term-type are found to be referentially unsuccessful over time, we can still explain why particular scientists or communities of scientists were engaging in successful practices by finding out the particular occasions on which they were grappling with something real.

Kitcher's main insight, I believe, is that referential success is sensitive to the target scientists have in mind: e.g., the underlying causal agent responsible for a set of observed phenomena or the fundamental properties of the entity posited by some theoretical account. From the perspective of Newtonian physics, for instance, Galileo's early tokens of '*gravitas*' failed to refer when he intended to use '*gravitas*' to refer to the intrinsic quality of bodies responsible for their tendency to move towards their natural place, namely, the center of the Earth in the case of terrestrial bodies.[8] In contrast, Galileo's early tokens of '*gravitas*' successfully referred to the approximately constant Newtonian force between the Earth and massive terrestrial bodies when his dominant intention was simply to use '*gravitas*' to refer to the causal agent responsible for the free fall of bodies near the surface of the Earth.

Is this insight sufficient to solve the issues we are concerned with? While there is no issue about its suitability as a theory of reference for scientific terms *per se*, Kitcher's theory does not appear to help selective realists establish a reliable connection between the achievements of a theory and the referential success of its central theoretical terms.[9] In my view, the most serious issue arises from Kitcher's focus on individuals. Why should particular scientists or, worse, particular historians of

[8]This seems to be the consensus on Galileo's theoretical understanding of gravity in *De Motu* and *De Meccaniche* (e.g., Jammer, 1957, p. 97; Koyré, 1966; Westfall, 1971, chap. 1; Hooper, 1998, p. 153; Massimi, 2010). See Westfall (1971, esp. p. 17, p. 22) and Koyré (1966, p. 35) for the insistence on Galileo's focus on the non-Aristotelian notion of natural places as multiple centers of an ordered cosmos. For a specific example of Galileo's intention to use 'absolute heaviness' ('*gravità assoluta*') to refer to the intrinsic quality of bodies, see Galileo's explanation of the continuous transition of terrestrial bodies moving upwards and subsequently falling once the gravitational force exerted by the Earth overcomes the initial force impressed on them in Chap. 17 of *De Motu* (esp., Galilei, 1590, p. 81).

[9]Psillos (1997), Stanford (2003b, 2006), and McLeish (2005) also raise several criticisms against Kitcher's account, but I do not think that they make it untenable. (i) Psillos (1997, p. 269) and Stanford (2003b, p. 557; 2006, pp. 148-50) argue that since successful reference is presumably only established by means of causal link in the case of deeply mistaken theories, Kitcher's theory reduces to and therefore faces the same issues as the original causal theory of reference. In Kitcher's defense, we do not need to have the sharp division between causal and descriptive modes of reference which Psillos attributes to Kitcher's theory (see, e.g., Psillos, 1997, p. 261). It is even an advantage of his theory that it allows for purely causal, purely descriptive and, with the appropriate modifications, hybrid reference-fixing mechanisms. (ii) Psillos (1997, p. 262) and McLeish (2005) also press Kitcher on what McLeish (2005, p. 669) calls the "discrimination problem", i.e., the problem that we, as interpreters, cannot legitimately discriminate between referentially successful and unsuccessful past tokens by appealing to historical or modal facts, or by relying on our own semantic intuitions. In Kitcher's defense, again, it was never assumed that we have perfect access to past scientists' mental states, or that the interpretative task is easy or infallible. In general, we need to look at past scientists' writings, practices, reports, and the contexts of their particular activities in order to formulate the best explanation of their achievements. (iii) Psillos's criticism that Kitcher's strategy makes past scientists' inferential practice look too incoherent by the light of current standards is probably the most damaging one (see Psillos, 1997, pp. 265-6). But even then, we might grant that scientists are sometimes vulnerable to linguistic slippages and fail to realize that they talk about different things with the same word.

science have any authority over the referential success of the central theoretical terms of a theory? Even if we are able to mount the best explanation for scientists' particular referential practices, we still face the issue that scientists often have widely different beliefs, attitudes and methodologies, even when they work with the same theory or on the same topic. For instance, the late Galileo was careful enough not to commit to any fine-grained and fundamental causal explanation of gravity in contrast to Descartes who, ironically, criticized him on this point (e.g., Descartes, 1991, p. 124 [AT II: 380]). The empiricism and caution of the late Galileo presumably made him more referentially successful than Descartes. But what does it have to do with the achievements of Galilean and Cartesian physics? Koyré (1965, p. 186, 1966, part III) and Gabbey (1998, p. 666) even point out that Galileo's reluctance to speculate about the fundamental nature of gravity is likely to have led him to miss the exact formulation of the inertial principle, and that Descartes's mistaken views about gravity were precisely what led him to be the first to successfully formulate this principle. The particular paths that individual scientists take to reach successful results do not seem to be reliably correlated to their ability to pick out real entities. Of course, the claim is not that the correlation is always unreliable but rather that the focus on scientists' idiosyncratic uses of term-types is too fine-grained to enable selective realists to reliably assess whether the central terms of successful theories latch onto the world or not. And even if we grant that scientists sometimes refer non-uniformly, what really seems to separate successful from unsuccessful tokens is the ability of a scientist to describe some causal agent at the right level of description and in the appropriate physical circumstances (see sections 4.3 and 4.4 below).

Turning to Psillos (1999, 2012), his solution to the combination problem is to take both descriptive and causal elements of reference-fixing to play an essential role for referential success. His theory of reference can be reconstructed as follows. A theoretical term t in a theory T refers to an entity x under three conditions:

(C) <u>Causal link</u>: t is introduced to pick out some causal origin x of a set of observed phenomena ϕ;

(S) <u>Satisfaction link</u>: x satisfies the core causal-explanatory description of ϕ associated with t in the theory T;

(T) Tracking condition: The core causal-explanatory description of ϕ captures the set of kind-constitutive properties of x that play an indispensable role with respect to T in the causal explanation of ϕ.[10]

In my view, the most interesting aspect of Psillos's theory is its robust and precise identification of the shared burden of reference-fixing via causal contact and via satisfaction. We can always stipulate that a term in a theory picks out the causal origin of some observed phenomena and thereby ensure the existence of some referent. But if the theory does not do any work, as it were, we do not seem to have any means to ensure that the term does not pick out multiple entities. We need, in general, some reasonable amount of information in order to uniquely circumscribe the referent of a term. Inversely, if we do not link successive theories to the causal origin of the phenomena they are supposed to account for, we undermine their ability to talk about the same entity if they say different things about it. Psillos's theory can be seen as avoiding these two issues as follows: the term is first linked by causal contact to a set of referents $\{r_1, ..., r_n\}$, with $n \geq 1$; the core causal-explanatory description either fails to pick out any of these referents or selects a subset of them $\{r_1, ..., r_m\}$, with $m \leq n$; in principle, the tracking condition ensures that there is only one referent left in $\{r_1, ..., r_m\}$ which has the required properties.

For instance, 'heaviness' in Galileo's later works is introduced to pick out some causal origin for the observed free fall of terrestrial bodies. Because Galileo takes heaviness to be a coarse-grained quality of bodies without committing to its deep nature or metaphysical status, Newton's gravitational force does satisfy the phenomenological descriptions associated with 'heaviness' (e.g., "gravity is responsible for the differences in velocities between the early and later times of the free fall of a terrestrial body").[11] Yet, these descriptions do not capture the kind-constitutive

[10]By 'core causal-explanatory description' Psillos means the description of x that anything has to satisfy in order to play the same causal role as x with respect to ϕ. By 'indispensable' he means that the kind-constitutive properties cannot be replaced by other non-*ad hoc* properties in T playing the same role in the causal explanation of ϕ (see Psillos, 1999, p. 110).

[11]Here and after, I will follow Koyré (1966, part III, sec. 3) and understand Galileo's notion of gravity in his later works as a macroscopic "empirical" property of massive bodies.

properties of Newton's gravitational force, say, its dependency on the square of the inverse distance between massive bodies and on their respective masses, and Psillos's theory thus entails that 'heaviness' does not refer to Newton's gravitational force.

How then does Psillos's theory fare with respect to the problem of referential failure and Stanford's challenge? We are confronted with two main issues here. First, Psillos's appeal to kind-constitutive properties appears to be overly restrictive. We might take some kind-constitutive properties to play an indispensable role in the causal explanation of some observed phenomena. And yet there might not be any causal agent possessing these properties or any well-delineated kind of entity associated with these properties. Psillos (2012, p. 226) addresses this issue and suggests replacing kind-constitutive properties by "stable identifying properties", provided that they take an indispensable part in the causal explanation of the observed phenomena with respect to the theory of interest. Yet, Psillos does not provide much detail about the notion of "stability" (Psillos, 2012, pp. 224-7). Equating stable properties with properties conserved under theory-change begs the question (see, e.g., Stanford, 2003a,b, 2006; Chakravartty, 2007, p. 46; Ladyman et al., 2007, p. 89). Appealing to properties playing an indispensable role in the explanation of the observed phenomena ϕ does not help either. For if a theory is superseded by a more successful one in the same domain, these properties will remain indispensable only if they appear to be so by the light of the successor theory. Likewise, I agree with Stanford's criticism that Psillos's earlier appeal to scientists' judgments does not work (see Psillos, 1999, pp. 112-3; Stanford, 2006, sec. 7.3; Ladyman et al., 2007, pp. 90-1). Past scientists might be deeply mistaken about the properties they deem stable and, more generally, this solution would bring us back to the same sorts of issues that beset Kitcher's theory.

Second, and more importantly, Psillos's theory makes referential success both too easy and too hard to achieve. His own example of the luminiferous ether and the classical electromagnetic field illustrates well the first case. Even if we accept that these entities play the same causal role (e.g., dynamical structure for the propagation of light waves) and share a set of stable core causal properties (e.g., continuous medium, repository of the kinetic and dynamical energy of light), this does not seem to be sufficient for granting that 'luminiferous ether' refers to the classical electromagnetic field. As French (2014, pp. 4-5, p. 125) rightly points out, some of the core individuating properties of the luminiferous ether such as its mechanical nature and its molecular constitution are not shared by the classical electromagnetic field. If we eliminate these properties as parts of what fixes the referent of 'luminiferous ether', referential success is achieved at the expense of replacing (as it were) the ether as a self-standing entity with clear identity conditions by a small cluster of stable properties. In this case, however, we face again the issue of radical referential indeterminacy, i.e., the issue that 'luminiferous ether' might refer to radically different types of entities, including an empty space containing collections of photons and other sorts of particles.

Psillos's theory also makes referential success too hard to achieve: the same theoretical term might be associated with radically incompatible core causal descriptions in two different theories and still successfully refer in both cases if the descriptions of each theory are used at the appropriate level. To give one striking example: (i) the gravitational force in classical Newtonian mechanics essentially plays the same causal role as the curvature of space-time in classical General Relativity with respect to terrestrial gravitational phenomena; (ii) the term 'gravity' is associated with radically incompatible core causal-explanatory descriptions in the two theories (e.g., gravitational forces are non-local while gravitational effects propagate locally in standard curved space-times); (iii) and yet, near the Earth and more generally in contexts where the curvature of space-time is sufficiently small and the observational time scale is sufficiently large, we seem to be justified in identifying the causal origin of gravitational effects with an instantaneous force between massive bodies.

These last two points signal that Psillos's theory of reference does not have appropriate resources to address what might be called the "problem of referential tracking": given some theory, at which level of description should we locate the causal agent(s) responsible for a set of observed phenomena? Consider for instance Galileo's mature account of terrestrial gravitational phenomena. Should we restrict our focus to medium-size entities close to the surface of the Earth, ignore the Earth itself, and take the set of macroscopic terrestrial properties to be the appropriate locus of reference? Should we “zoom out", include the Earth, and take ‘heaviness’ to refer to a force relating massive bodies? Should we “zoom out" even more and take ‘heaviness’ to refer to the smoothly curved structure of space-time? Or perhaps should we rather “zoom in" and take ‘heaviness’ to refer to collections of gravitons? The problem of referential tracking, in other words, is to circumscribe some unique referent given equally plausible candidates specified at different levels in the causal structure underlying a set of observed phenomena. I will argue in the next two sections that the most epistemically reliable way of addressing this problem is to focus our attention on the range of parameters over which the theory is empirically reliable, and I will adjust Psillos's theory of reference accordingly.[12]

4.3 Reference and Effective Theories

> Following Galileo's lead of ignoring complications and searching for simplicity, physicists have developed formalisms in which the separation between different ways of talking—“effective field theories"—is precise and well-defined. (Carroll, 2016, p. 113)

Despite its deficiencies, Psillos's theory has the merit of offering a convincing solution to the combination problem and shifting the original problem of referential failure to the more tractable issues of referential tracking and stability. In this section, I will propose a theory of reference which addresses the issue of referential tracking by drawing on the framework of effective theories. The idea, in a nutshell, is to identify the limited objective context where the theory is empirically reliable and restrict the potential referents of its central terms accordingly. I will discuss the issue of stability and Stanford's challenge in the next section.

Recall that the problem of referential tracking consists in selecting some appropriate causal agent(s) for a set of observed phenomena given equally plausible candidates specified at different

[12] For other extensions of Kitcher's and Psillos's proposals and distinct proposals which I will not discuss for lack of space, see, e.g., Chakravartty (1998, 2007); Saatsi (2005); Field (1973), McLeish (2006), and Landig (2014).

levels in the causal structure underlying these phenomena (or in the causal chain leading to these phenomena). As we saw above, Psillos's appeal to "kind-constitutive" or "stable" properties is far from ensuring that a term tracks the appropriate causal agent. There might be no causal agent possessing the required set of kind-constitutive properties or radically different kinds of causal agents possessing the required set of stable properties. Kitcher's theory, by contrast, appears at first sight to offer a better solution: a scientist successfully isolates the appropriate causal agent if she intends to refer to this causal agent. Yet we still face a serious issue here. The scientist needs to extract some information from a given theory in order to circumscribe potential candidates and select the appropriate target in her mind, and the selection process is likely to depend significantly on how reliable she thinks that information is and therefore vary significantly from individual to individual.

How can we constrain referential tracking without relying on scientists' particular beliefs then? Suppose that we restrict the set of appropriate causal agents for a given set of observed phenomena by means of the physical context where the theory is putatively applicable. For instance, Galileo's mature law of free fall applies to the idealized motion of bodies falling in a hypothetic vacuum near the Earth, but it does not apply to the motion of celestial bodies because Galileo's account presupposes that they follow circular and unconstrained uniform trajectories.[13] Hence, Galileo's term 'heaviness' tracks entities in the vicinity of the Earth but not beyond. Does it mean that the term tracks collections of gravitons near the surface of the Earth? We need to be careful here. The putative domain of applicability of a theory does not only depend on its internal principles and constraints but also on how we intend to define the scope of the theory in the first place. If we take the theory to give a fundamental and complete description of the causal origin of some phenomenon, we need to consider the whole world as the appropriate locus of reference for the terms of the theory. If we take the theory to give only a phenomenological description of this causal origin (e.g., "gravity causes macroscopic bodies to fall downwards near the surface of the Earth"), we need to restrict ourselves to the domain of macroscopic entities. Either way, the selection

[13]For a discussion about Galileo's inertial principle, see Drake (1964), Koyré (1966, part III), and Hooper (1998).

process is still overly sensitive to particular interpretative choices, which brings us back again to the set of issues that beset Kitcher's theory.

One solution is to specify the physical context at work in referential tracking by means of the limited range over which the theory is empirically reliable and impose the following semantic constraint: a term t in a theory T tracks some entity x only if x is located at a level and in circumstances specified by the range of scales over which T is empirically reliable. How do we determine this range if we are to evaluate referential success before theory-change? In general, this range simply corresponds to the range over which the theory has been *found* to be empirically accurate. In some cases, however, we may be able to estimate the limited range over which a theory is likely to remain empirically reliable even if we have not yet probed phenomena at the relevant scales in experiments. As we will see below, effective theories provide a paradigmatic example. We might also be able to extend this range if we find that the theory makes accurate predictions in new regimes. Be that as it may, the important point for now is that scientists' and interpreters' dominant intentions and expectations do not significantly interfere at least with the specification of the range where the theory has been found to be empirically accurate. They certainly need to pick a reasonable standard of measurement accuracy and the extent of this range depends on the experimental achievements reached at a certain time. But apart from that, we only need to assume that the experimental predictions derived from the theory depend on, or can be associated with, a set of independent parameters and that, as we vary these parameters, the comparison of predictions with empirical data determines the limited range over which the theory is empirically accurate. We can thus use this sort of objective limit to adjust the semantics of physical theories and assess the referential success of their theoretical terms at a given time without facing the issues discussed above.

Two comments are in order. First, the new semantic constraint provides a solution to the problem of referential tracking in the sense that it restricts the set of appropriate causal agents responsible for some observed phenomena. As we will see in section 4.4, the main reason for making this semantic adjustment is epistemic, i.e., a causal agent is deemed "appropriate" if we have reliable epistemic access to it. Second, an objective context in theoretical physics simply amounts to a set of physical conditions specified by a range of physical scales. As I will explain in section 4.4, the specification of an objective context in less fundamental areas of scientific inquiry is less straightforward and requires, in particular, background assumptions about the set of entities, properties, and relations which characterize the target system.

Now, this solution to the problem of referential tracking is compatible with Psillos's solution to the combination problem only if it is possible to separate the descriptions of the theory according to its empirical limitations. This is where the framework of effective theories proves to be particularly useful, and there are two general features which are important here: (i) the structure of an empirically successful effective theory incorporates a robust specification of the scales at which it is likely to be empirically inaccurate before we probe these scales in experiments; (ii) its structure is such that we can separate its descriptions into two sets according to these empirical limitations. I will provide concrete examples below. For now, we can understand these two general features as follows (more details are given in chapter 3).

Suppose that the expressions of a theory T, say, its dynamical equation, depend on some parameter E, with $0 \leq E < +\infty$. Then T is an "effective theory" if it satisfies the four following properties:

1. The equation depends on some non-trivial scale Λ, with $0 < \Lambda < +\infty$;

2. It is possible to include increasingly many terms depending on E/Λ into this equation which are consistent with the core principles governing its structure, with one arbitrary coefficient for each new term introduced;

3. These terms are systematically organized according to the importance of their contributions to predictions below and above Λ;

4. As we include increasingly many terms of increasing order in E/Λ, the predictions derived from the equation remain approximately the same, say, for $E \ll \Lambda$ and become increasingly large for $E \geq \Lambda$ (and infinite if we include an infinite number of terms).

The interpretation of Λ as an intrinsic empirical limiting scale is first suggested by the fact that if we try to make T as empirically accurate as possible for $E \ll \Lambda$ by including increasingly many new terms, its predictions become ultimately deficient for $E \geq \Lambda$. This predictive pattern is, of course, not sufficient by itself. The interpretation is ultimately grounded in the experimental profile of theories displaying the same predictive pattern. As it turns out, existing cases of low-energy effective theories do become increasingly empirically inaccurate as we probe phenomena closer and closer to their limiting scale. Adding increasingly many terms to compensate for experimental discrepancies and maintaining their predictive power is usually not sufficient; and in the limit, if we were to include increasingly many terms, the resulting theories with an infinite number of terms would become predictively powerless at their limiting scale. We would not be able to collect an infinite number of empirical inputs in order to fix their free parameters. And we would not be able to select a finite number of terms in order to make approximate predictions since all the terms contribute equally to predictions at the limiting scale itself. The experimental profile of existing effective theories thus supports the interpretation of Λ as the maximal predictive limit of T. And the scale Λ, in turn, separates its descriptions $D(E)$ into two sets, i.e., $\{D(E), E < \Lambda\}$ and $\{D(E), E \geq \Lambda\}$.

Now, if we rely on these two general features of effective theories and follow Psillos's lead, we can formulate an objective context-dependent theory of reference (CST*) as follows. A term t in a theory T refers to an entity x under three conditions:

(C*) Causal link: t is introduced to pick out some causal origin x of a set of observed phenomena ϕ within the objective context $\mathcal{C}$ delineated by the empirical limitations of T;

(S*) Satisfaction link: x satisfies the core causal-explanatory description of ϕ associated with t in T;

(T*) Tracking condition: The core causal-explanatory description of ϕ ranges over, or is associated with, the objective context $\mathcal{C}$.

To see how (CST*) works as a semantic constraint with toy-models of effective theories, let us first look at the standard Galilean and Newtonian laws of free fall, rewritten in their mathematically most simple modern formulation for conceptual clarity (see Table 4.1 below). The target system in the Galilean case is a heavy body dropped at some height $z(t)$ from the ground. The target system in the Newtonian case is a body of mass m located at some distance $r(t)$ from the center of the Earth. In each case, the equation of motion is derived from the action S defined as the integral over time of the Lagrangian L, where L encodes information about the dynamics of the system. G is the universal gravitational constant, M the mass of the Earth, and c some arbitrary constant.[14]

Galilean	**Newtonian**
$S_G = \int dt L_G(z(t))$ $L_G = \frac{m}{2}(\frac{dz}{dt})^2 - mgz$	$S_N = \int dt L_N(r(t))$ $L_N = \frac{m}{2}(\frac{dr}{dt})^2 - \frac{GMm}{r}$
$\frac{d^2z}{dt^2} = -g$ $g =$ constant Translation invariance ($z \to z + c$)	$m\frac{d^2r}{dt^2} = -mg(r)$ $g(r) = \frac{GM}{r^2}$ No translation invariance ($r \to r + c$)
g: heaviness of matter in a vacuum (universal quality of terrestrial bodies, local internal action).	$g(r)$: interaction force exerted by the Earth on the body per unit mass (relational property, action at a distance)

Table 4.1: The Galilean and Newtonian laws of free fall.

How can we construct effective versions of the Galilean and Newtonian laws of free fall? As it turns out, we already know their closest successor, i.e., respectively, Newton's theory and classical General Relativity. So we can simply follow the "top-down" strategy by appropriately restricting the range of the more comprehensive theory and eliminating its theoretical constituents which do not contribute significantly to predictions within this range. In the Galilean case, for instance, we can replace the parameter $r(t)$ by $z(t) + R_E$ in the Newtonian equation of motion, with R_E the

[14] I eliminated the mass of the system in the Galilean equation of motion to avoid attributing to Galileo a distinction between the mass and the gravitational quality of heavy bodies, but kept it in the Lagrangian for dimensional consistency.

radius of the Earth, assume $z(t) \ll R_E$, and expand $g(r)$ in terms of $z(t)/R_E$ (see Eq. 4.1 below).

We can also pretend that we do not yet know the more comprehensive theory, i.e., pretend that we are dealing with the theory at the time it is still a live concern, and follow the “bottom-up" strategy. We first identify a limited range where we think that the theory is reliable. For instance, we may have found that the Galilean law makes slightly inaccurate predictions for heavy bodies dropped too far from the ground. Or we may suspect from the infinite value of $g(r) = GM/r^2$ in the limit $r \to 0$ that the Newtonian law is mathematically inadequate for describing arbitrarily small bodies moving arbitrarily close to one another. Then, we restrict the range of the theory by introducing some arbitrary limiting scale, namely, a large-distance scale z_0 in the Galilean case and a short-distance scale r_0 in the Newtonian case. And finally, we include all the possible terms depending on the limiting scale which are allowed by the symmetries of the original theory, with one arbitrary coefficient for each new term. As we perform these steps, we do not need to know anything about the value or the underlying meaning of the limiting scale, i.e., that z_0 turns out to be the radius of the Earth R_E and r_0 its Schwarzschild radius $2GM/c^2$, with c the speed of light. The value of these scales and of the additional coefficients is ultimately fixed by means of empirical inputs, at least for a finite number of them. In general, however, we need to assume that the dimensionless constants of the theory are of order 1 to obtain a first estimate of the value of the limiting scale (i.e., we need to take $a_i, b_i = O(1)$ in Eqs. 4.1 and 4.2 below).

Now, whether we follow the top-down or the bottom-up strategy, the effective Galilean and Newtonian laws take the following form:

<u>Effective Galilean law of free fall</u>:

For $z \ll z_0$,

$$\frac{d^2z}{dt^2} = -g\Big(1 + a_1\frac{z}{z_0} + a_2(\frac{z}{z_0})^2 + ...\Big) \tag{4.1}$$

Effective Newtonian law of free fall:

$$
\begin{aligned}
&\text{For } r \gg r_0, \\
&m\frac{d^2r}{dt^2} = -\frac{GMm}{r^2}\Big(1 + b_1\frac{r_0}{r} + b_2(\frac{r_0}{r})^2 + ...\Big)
\end{aligned} \tag{4.2}
$$

with a_i and b_i free parameters.[15] In each case, the most complete effective theory includes an infinite number of terms, i.e., all the possible terms allowed by the original dynamical variables and symmetries of the equations.[16] We may also need to introduce additional limiting scales and new terms depending on these scales, such as the characteristic scale $G\hbar/2c^3$ used for quantifying the importance of quantum gravitational effects (with $\hbar$ the reduced Planck constant).

These two simple examples show that the structure of an effective theory gives us precise constraints for implementing (CST*).[17] Consider the effective Newtonian law in Eq. 4.2 for instance. Suppose that we have empirical inputs to fix some of the parameters b_i and estimate the value of the limiting scale r_0, and that we find that the theory is empirically successful for $r \gg r_0$. The structure of the theory gives us good reasons to believe that its predictions will become unreliable when gravitation becomes too strong and when the interaction between two massive bodies occurs at too short distances. For the most complete effective theory, the separation is sharp and precise. The infinite expansion $1 + b_1 r_0/r + b_2(r_0/r)^2 + ...$ has an infinite value for $r_0/r \geq 1$ and becomes predictively powerless for $r \sim r_0$. For the effective theories defined by means of a finite number of terms, the separation is not as vivid and sharp. But as explained above, if we consider their predictive pattern and the experimental profile of existing theories which display the same pattern, we still have good reasons to think that their predictions are inaccurate for $r \ll r_0$. In both

[15] More details about the first order relativistic and quantum corrections to the non-relativistic gravitational potential can be found in Donoghue (1995), Burgess (2004), and Blanchet (2014).

[16] Sometimes, however, we need to break some of these symmetries: for instance, the effective version of the Galilean law of free fall breaks translation invariance along z.

[17] Of course, I do not mean to suggest with these examples that every physical theory can be formulated as an effective theory or that these simple Taylor expansions account for the entirety of gravitational phenomena in their domains. It is worth emphasizing here that we often need more than one effective theory within a given domain. Nuclear physics is a particularly good example in that respect (e.g., Chiral Perturbation Theory, Heavy Quark Physics, Non-relativistic Quantum Chromodynamics, etc.).

cases, and provided we have appropriate empirical inputs, the structure of the theory allows us to delineate the range where it is likely to remain empirically reliable and to separate its descriptions accordingly. And if we use these features to implement (CST*), we find that the term 'gravity' refers to a force at sufficiently large distance scales $r \gg r_0$ whose strength is given by Eq. 4.2 with $r \gg r_0$. I have, of course, not said anything yet about the issue of referential stability (cf. section 4.4). But this simple toy-model already shows that the structure of an effective theory gives us precise constraints for selecting entities at the appropriate level.

Compare with the original Galilean and Newtonian laws of free fall. On the face of it, these theories are putatively applicable across all scales (apart from the trivial scale $r = 0$ in the Newtonian case) and their structure does not encode information about some non-trivial scale at which their predictions are likely to break down. If the Newtonian theory were our most fundamental account of the phenomenon of free fall, we might wrongly take it to describe some fundamental kind of entity specified, say, at arbitrarily short distance scales. Now, of course, we may also have external reasons to believe that the theory is unreliable at short distances, decide to restrict the potential referents of the term 'gravity' accordingly, and use only some of the descriptions of the theory to select an appropriate set of referents. This selection is likely to be uncontroversial if we have found some empirically successful and more comprehensive theory (i.e., classical General Relativity in this case). At the time the theory is still a live concern, however, the selection is likely to be more controversial and significantly depend on particular interpretative choices. If we have appropriate empirical inputs at sufficiently large distances, the framework of effective theories allows us to obtain some estimate of the physical context where the theory is likely to remain empirically reliable, separate its descriptions accordingly, and select a subset of entities within a well-delineated physical context.

Before discussing the issue of stability and examining how (CST*) works for problematic historical cases, it will be instructive to first look at current theories given what we expect from future theories. A paradigmatic example is the possible existence of new space-time dimensions at short-distance scales in string theories. To see how this works, consider the following Kaluza-Klein toy-model of “dimensional reduction" from five to four space-time dimensions:

$$
\begin{aligned}
S &= \frac{1}{2}\int d^5x \partial_\mu\phi(x)\partial^\mu\phi(x) \\
&= \sum_{n=-\infty}^{n=+\infty}\frac{1}{2}\int d^4x_i(\partial_\mu\psi_n(x_i)\partial^\mu\psi_n(x_i) - m_n^2\psi_n^2(x_i))
\end{aligned}
\tag{4.3}
$$

$$
\text{with} \begin{cases} \phi(x_i, y) = \sum_{n=-\infty}^{n=+\infty}\psi_n(x_i)e^{iny/R} & (i = 1, ..., 4) \\ m_n^2 = \frac{n^2}{R^2} \end{cases}
$$

The first line gives the action S of a massless field living in a 5D space-time with one “compact" dimension—say, a circle of radius R, with $\phi(x)$ the field variable specifying the value of the field system at each point x of the 5D space-time. An observer living in this higher-dimensional space-time can move both along the four directions x_i of the 4D space-time and along the direction y of the circular dimension, with $x = (x_i, y)$. The second line is obtained by eliminating the y-dependent terms in Eq. 4.3 after separating the x_i- and y-dependent components of the original field $\phi(x_i, y)$. The resulting action S describes an infinite number of “new" fields $\psi_n(x_i)$ of increasing mass m_n living in the 4D space-time, with one massless field $\psi_0(x_i)$.[18]

So far, I have described the “top-down" derivation of a 5D theory into a 4D theory. But we can also take the “bottom-up" perspective of current experimenters probing shorter and shorter distance scales (or, equivalently, higher and higher energy scales). Assuming that R is very small, an observer living in the 5D world specified by the action S has good reasons to believe that the world is four-dimensional at large enough distance scales $L \gg R$, i.e., at a level and in circumstances where she does not detect the effects of the fifth circular dimension. According to the “effective" 4D description of this world, however, she will eventually detect increasingly many new types of particles (i.e., field patterns) with increasingly heavy masses n^2/R^2 if she probes shorter and shorter distance scales. Knowing the mechanism of dimensional reduction, she will have good reasons to believe that these new types of particles stand for the effect of a fifth dimension and

[18] For introductions to the topic of extra dimensions, see, e.g., Csaki (2005), Perez-Lorenzana (2005), and Sundrum (2005).

therefore that she lives in a 5D space-time. Seen in their entirety, the 4D and 5D worlds look radically different, both in terms of particle content and topological structure. And yet, at large distances, the observer appears to be justified in taking '$\psi_0(x)$' to refer to a massless field living in a 4D space-time relative to these scales even after having discovered the new types of particles. At much shorter distance scales, the observer might detect effects indicating the presence of a new circular dimension of radius $R' \ll R$ and realize again that she was justified in taking '$\phi(x)$' to refer to a massless field living in a 5D space-time only relative to sufficiently large distance scales. Whether the observer is ultimately justified in holding these beliefs, the important point here is that it is crucial to specify some limited physical context if we want the terms of the theory to pick out determinate entities at a particular level. (CST*) simply provides the semantic adjustment required to deal with those sorts of situations.

4.4 Stability and Objective Context-Dependence

We have seen that we can constrain the terms of a theory to pick out entities at a particular level of description by identifying the objective context delineated by the empirical limitations of the theory. In this section, I will argue that this sort of constraint allows us to solve the issue of referential stability, i.e., the level of description is "appropriate" in the sense that the terms of the theory are likely to remain referentially successful under theory-change at this level. I will also further clarify the notion of objective context and apply this strategy to the problematic cases of the phlogiston theory and 19th century theories of the luminiferous ether.

Note, first, that (CST*) reduces the risk of referential failure. The case of effective theories is straightforward. The structure of an effective theory delineates the scales where it is likely to make inaccurate predictions and thereby provide false information about its target system. The theory therefore gives us good reasons to believe that its descriptions ranging over these scales are false and fail to be satisfied by anything real. By imposing the tracking condition (T*), we thus have a direct way to ensure that these descriptions do not play any role in reference-fixing, i.e., to reduce the risk that the terms of the theory fail to refer to any of the candidates picked out by causal contact as specified by the condition (C*).

In the general case, if we are able to associate the descriptions of a theory with different physical contexts, (CST*) still implies that the central terms of the theory do not refer to entities located at a level and in circumstances which have not been put to the test yet, i.e., where we do not yet have any reason to trust the content of the theory. Suppose for instance that we are working with some empirically successful theory and that, in contrast to effective theories, its structure does not contain any intrinsic empirical limitation. If we take the theory by itself, it may either continue to make accurate predictions or break down in new regimes. In this case, a reasonable option would be to remain agnostic about the referential success of the terms of the theory which are supposed to pick out entities in the corresponding domain. It does not mean, however, that we should remain agnostic about the referential success of the terms which are supposed to pick out entities in the domains where the theory has been found to be empirically accurate. (CST*) therefore enjoins us to err on the side of caution: we should look for entities in the domains where the theory has been put to the test and, if we are to assess the referential success of its central terms at a given time, assume that they fail to refer to anything real in unexplored domains until there is evidence to the contrary.

As it turns out, there are also good reasons to believe that the terms selected through (CST*) will *remain* referentially successful under theory-change. Again, the case of effective theories is special and I need another one of their distinctive features which I have not yet discussed to make this point: namely, that the descriptions which are the most relevant for predictions within $E \ll \Lambda$ are largely insensitive to the descriptions which are the most relevant within $E \gg \Lambda$.[19] Consider for instance the effective Newtonian law of free fall with only a few first order terms and suppose that its predictions have been found to be accurate at large distances $r \gg r_0$. If we discover a radically new theory which reveals that these predictions are slightly inaccurate, we

[19]For a similar argument employing renormalization group and effective field theory methods in the specific context of QFT, see J. D. Fraser (2018) and Williams (2019b). For a discussion about the link between naturalness and inter-scale insensitivity, see, e.g., Williams (2015). Note that a theory might be unnatural (in the sense that some of its free dimensionless parameters are not of order one) and still contain parts which are largely insensitive to potentially new types of high-energy physics (see, e.g., Donoghue, 2020, p. 4, for a similar point). We just need to be more selective in this case.

can always add higher-order terms in the effective Newtonian theory and fix their coefficients with empirical inputs at large distances in order to compensate for the predictive discrepancy. This move is, of course, largely *ad hoc*. But it shows that the higher-order terms can be used to encode the contributions of new physics at large distances according to their relevance and that these terms therefore do not simply correspond to arbitrary modifications of the theory with no physical significance whatsoever. The ability of higher-order terms to stand for fine-grained features of new physics is also supported by explicit derivations of effective theories from more comprehensive ones, as it is for instance the case with the Galilean and Newtonian laws of free fall. And, in general, the structure of an effective theory is such that we can parametrize the contributions of any type of new physics at large distances up to an arbitrarily high degree of precision by adding increasingly many terms depending only on the degrees of freedom of the original theory. In the Newtonian case, we can even include such terms by preserving *all* the core principles of the original theory (e.g., the structure of the classical Newtonian background space-time).

Now, the crucial point is that the contributions of the higher-order terms become increasingly negligible at large distances $r \gg r_0$, no matter what the new physics looks like. And insofar as these higher-order terms stand for fine-grained features of new physics, this shows that the descriptions of the effective theory which are relevant at large distances are largely insensitive to the details of this new physics. It affects at most the value of the parameters of the first few order terms. Hence, in the case of effective theories at least, the tracking condition (T^*) selects core properties at a particular level which, in general, do not significantly depend on more fine-grained features (or coarse-grained ones). This, in turn, gives good reasons to believe that the central terms of an effective theory selected through (CST^*) successfully pick out determinate entities no matter what they will look like from the perspective of a future theory.

In the general case, the principle of selective reference at work in (CST^*) still gives us a reliable way of identifying the terms in the theory which are the *most* likely to remain referentially successful under theory-change. Suppose again that we are able to delineate the limited context where the theory is empirically reliable (e.g., macroscopic distance scales). Suppose further that we can separate the core causal-explanatory descriptions of the theory and thus different causal components entering into the causal explanation of phenomena according to this context (e.g., the mass of a macroscopic body and the set of microscopic particles which constitute it in order to explain its trajectory over sufficiently long periods of time). Now, unless there is evidence to the contrary, it is reasonable to attribute a greater degree of confirmation to those descriptions of causal components which participate more directly in the explanation of the phenomena according to this context (i.e., the mass of the macroscopic body given different types of trajectories and constraints) and do not significantly depend on causal components participating less directly in this explanation (i.e., the microscopic constitution of the body). Moreover, the empirical limitations of the theory also give us good reasons to adopt a differentiated epistemic attitude towards these descriptions. If we have not yet probed the system at some level (e.g., at short-distance scales), we do not yet have any good reason to expect that the predictions of the theory will remain accurate at this level and therefore that the theory describes well the causal components entering into the explanation of phenomena at that level. By contrast, if we have probed the system at some level and found the theory to be empirically accurate (e.g., at macroscopic scales), we do have good reasons to believe that the theory describes well the causal structure of phenomena at this level. And overall, then, we have better reasons to believe that the characterization of entities at this level will be (approximately) retained in future theories and therefore that the terms picking out entities specified at that level will remain referentially successful under theory-change.

One might still be worried that (CST*) does not handle well cases of radical ontological discontinuities. Consider again Galileo's and Newton's laws and assume for the sake of the argument that Galileo's mature description of gravity as an intrinsic coarse-grained quality of bodies is radically incompatible with Newton's description of gravity as a relation between massive bodies, even in the limited context where both Galileo's and Newton's laws are empirically reliable. (CST*) yields the following results. In the Galilean case, (C*) links 'gravity' to the set of terrestrial causal agents responsible for gravitational effects near the Earth; (S*) restricts the set of causal agents to those which possess a specific set of properties described by the theory; (T*) restricts this set of properties to those which can be associated with terrestrial macroscopic contexts. In principle, then, (CST*) isolates heaviness as the only referent of ‘gravity’ in terrestrial macroscopic contexts. In the Newtonian case, a similar story applies and, in principle, (CST*) isolates the gravitational force as the only referent of ‘gravity’ in terrestrial macroscopic contexts.

We have two choices here. (i) Either we grant that (CST*) fails in some cases: e.g., the term ‘gravity’ or ‘heaviness’ in Galilean physics does not refer to anything real by the light of Newtonian physics. I suspect that most cases of referential failures concern domains which have not been put to the test yet and therefore that this solution is acceptable. I will give some evidence below with ‘phlogiston’ and ‘luminiferous ether’ and, in the present case, it does seem plausible to reinterpret the gravitational force $mg(z) = mGM/(z + R_E) \sim mGM/R_E$ as a macroscopic property of terrestrial bodies close enough to the surface of the Earth ($z \ll R_E$). (ii) Or we maintain that (CST*) works in all cases and regard problematic cases to be problematic only insofar as they enter in conflict with some further assumption about ontological reduction. That is, in the previous example, (CST*) implies that both ‘heaviness’ and ‘gravitational force’ successfully refer, but it does not address the further question of whether the referent is the same or not, and therefore allows for referential success, stability, and discontinuity to be compatible with one another.

I am inclined to follow this second route for two main reasons. First, it is an advantage of a theory of reference that it does not smuggle in too many metaphysical assumptions. Second, the empirical success of a theory and the robustness of its descriptions within a limited context still give us independent grounds to believe that the central terms of the theory restricted to this context are referentially successful before and after the advent of a new theory, especially in the case of effective theories.

Consider again the case of extra-dimensions for instance. The scientist who lives in a 5D space-time appears to be justified in believing that the world is four-dimensional and contains one kind of massless particle at large enough distance scales, even after the discovery of new data supporting the existence of a new dimension, precisely because the descriptions of the 4D theory are empirically well-supported at large distances and can be shown to be largely independent of the effects of the fifth circular dimension at these distances (i.e., the effect of arbitrarily massive particles is negligible at large distances). Accordingly, (CST*) requires us to make the following semantic adjustment: at large distances, the terms 'space-time' and '$\psi_0(x_i)$' in the 4D theory refer, respectively, to a 4D space-time and a massless field living in four dimensions. And this adjustment does not introduce any issue of incommensurability: it is perfectly possible to compare the properties and the causal role that the two space-times and the different kinds of particles play at different levels and evaluate whether we are justified in identifying the ones with the others. This further task, however, is a matter of ontological reduction and not referential success, strictly speaking.

To summarize, the theory of reference (CST*) has the following advantages compared to Kitcher's and Psillos's theories. First, the theory does not rely significantly on scientists' dominant intentions or practices. Second, the theory does not rely on theoretical constituents which appear to be indispensable for the derivation of successful predictions but which do not characterize the main target of the theory at the relevant scale or even fall within its empirical reach (e.g., the continuum structure of space and time in non-relativistic classical gravitational theories). Third, the theory gives a precise characterization of Psillos's notion of "stable identifying properties" by reducing them to the properties which can be associated with the range of scales where the theory is empirically reliable and robust with respect to new physics or irrelevant causal agents. Fourth, the theory addresses Stanford's challenge by providing a reliable and principled way of assessing the putative referential success of the central terms of a theory before theory-change. As the example of effective theories shows, we do not even need to have any knowledge about the theoretical content of future theories if we want to estimate the scales where current theories are likely to remain empirically reliable. And, of course, we can find the empirical limitations of a theory by confronting it with data at the relevant scales without having found a better theory yet. Finally, the theory does not offer a "pyrrhic victory" to scientific realism (Stanford, 2003b, 2006). All we need is to be able to understand and formulate theories as effective theories or, if this proves impossible, to have some way of separating their descriptions according to their empirical limitations. And, in principle at least, these limitations leave ample space for the terms of the theory to pick out unobservable entities in some limited part of the world.

One important difficulty not addressed so far lies in the interpretation of the notion of "objective context". We might ask: in what sense does a range of physical parameters define an "objective context"? In what sense is this context "objective"? And what does it mean for reference-fixing to be indexed to such context as expressed by the condition (C*)? At least as far as we are concerned with physics, my proposal is that we should take an objective context to be nothing less and nothing more than the set of physical conditions specified by a range of physical scales: for instance, low temperatures and large mass densities. I intend this in a qualified operationalist sense. This range is directly determined by means of experimental procedures and perhaps in some cases also indirectly by means of previously tested theoretical relations. In high energy physics, for instance, the characteristic energy scales of the particles colliding in scattering experiments determine the characteristic distance scales of the interaction process. The former is operationally defined by the acceleration process of the incoming particles while the later is inferred from the former by assuming that the Einstein-Planck and the de Broglie relations hold at these energy scales. In both cases, we do not need to make any substantive assumption about the nature of space-time or quantum fields. Hence, by staying as closely operationalist as possible, we avoid as much as possible interpretative disagreements about the set of entities, properties, and relations which characterize a specific physical context.

Now, the advantage of physics is that the empirical limitations of theories are in general specifiable by means of a few independent parameters. In the case of high energy physics, we typically just need some energy parameter E. In the simplest case, then, the condition (C*) constrains reference to be fixed relative to a particular range of energy scales $E \in [\Lambda_1, \Lambda_2]$ in the sense that the terms of an empirically successful theory only pick out entities located at the level delineated by the energy scales Λ_1 and Λ_2. The conditions (S*) and (T*) constrain the entities selected to satisfy the descriptions $D(E)$ of the theory ranging over the same range of energy scales, and these descriptions attribute properties to these entities *at* a specific level.

We surely need a more metaphysically involved notion of objective context as we move towards applied physics and the special sciences: for instance, by endorsing background assumptions about the set of entities, properties, and relations that characterize the target system. In this case, the notion of objective context is more akin to a “window" into the target system, as it were, and the scope of this window, i.e., the set of elements involved in the characterization of the target system, can be restricted by using the empirical limitations of the theory. In the case of Galileo's law of free fall, for instance, we need to assume that the heavy body is falling towards the ground near the surface of the Earth and have some pre-theoretical understanding of what this involves. As we move towards the special sciences, we probably need to specify a variety of quantifiable parameters (e.g., population number, decay rate, etc.), as well as the experimental set-up, the materials involved, environmental conditions, and other non-quantifiable causal factors. The fact that the list goes on presumably makes it more difficult to identify stable core causal-explanatory descriptions in those areas.

Nevertheless, (CST*) still helps us to evaluate the most problematic cases of referential failure discussed in the literature (e.g., Kitcher, 1978, 1993; Laudan, 1981; Psillos, 1999; Saatsi, 2005; Chakravartty, 2007; Ladyman, 2011). Consider first the case of phlogiston.[20] Among the core properties required to identify the putative referent of ‘phlogiston', one finds that phlogiston is contained within different types of substances, including combustible ones, and released, in particular, during combustion and calcination processes. However, at the time the phlogiston theory was a going concern, namely, before the new oxygen theory of chemistry developed by Lavoisier became increasingly popular by the end of the 18th century, the phlogiston theorists did not have any good experimental constraints to further specify the exchange process and the substance(s) exchanged during combustion and calcination. They had clear evidence that something was exchanged. But this was not sufficient for taking the term ‘phlogiston' to refer insofar as some of its core properties were specified at a too “fine-grained" level of the exchange process for which the phlogiston theory was not shown to be empirically reliable. For instance, the experiments per-

[20]See, e.g., White (1932) and Siegfried (2002) for more historical details.

formed on metals and sulfur did not have the means to discriminate whether something was emitted or absorbed during a calcination or combustion process. Similarly, it was common to assume that phlogiston was lighter than air in order to explain the typical increase in weight of metals after calcination. The phlogiston theorists, however, did not have the experimental means of testing this assumption.

By studying a larger range of processes involving different types of substances, gases and water in a wider variety of experimental situations, such as sulfur and phosphorus in enclosed environments involving a limited amount of air and water, the advocates of the new oxygen theory could get a firmer experimental hold on the correct locus of the substance(s) exchanged during combustion, calcination, respiration, and other processes.[21] After all, if one believes that the principle of conservation of mass universally applies to chemical reactions, there is strong evidence that a gain of mass in most metals after calcination arises because of the participation of an external substance. Here the context is delineated, in particular, by the set of different types, volumes, and weights of the substances involved in distinct chemical processes.

Consider now the case of the luminiferous ether.[22] The theoretical term 'luminiferous ether' fails to refer because some of the core properties of the luminiferous ether, e.g., that the ether has a molecular structure with fine particles and that light waves are continuously transmitted by means of the mechanical action of this molecular structure, could not be specified in domains where 19th century scientists had a good experimental access. Their ability to successfully refer was indeed constrained by the objective context delineated in terms of distance scales. While there was no evidence for the fine-grained structure of light and its propagation at short-distance scales, there was strong evidence from observed diffraction and interference patterns that light had a wave-like structure. It was acceptable to speak about light waves, oscillating wave-like patterns and the like (which explains the empirical and explanatory success of 19th century theories of the luminiferous ether). But *contra* Psillos, there was not enough ground and there is not by our current light to

[21] This is, of course, not to say that all the terms of Lavoisier's theory are referentially successful (e.g., 'caloric') or that Lavoisier's theory was more empirically adequate than the phlogiston theory at that time. See Chang (2010) for a recent re-evaluation of the Chemical Revolution.

[22] See, e.g., Schaffner (1972) and Darrigol (2000) for more historical details.

take ‘luminiferous ether’ to refer to anything real, and even less so to the classical electromagnetic field. Relative to large-distance scales, we are not justified in believing that light waves propagate thanks to the mechanical action of an entity possessing a molecular structure (compared, say, to the “apparent" instantaneous propagation of gravitational effects for weak gravity and large observational time scales).

4.5 Objections and Replies

In response to the theory of reference presented in section 4.3, one might either reject the relevance of the focus on reference for defending scientific realism or agree with its relevance but deny that the theory is correct. This goal of this section is to briefly address various forms of these two types of objections. I will leave aside structuralist complaints since I am concerned here with the standard form of scientific realism (see, e.g., Worrall, 1989; Saunders, 1993; Ladyman, 1998, 2011; Ladyman, 2011; French, 2014).

Objection 1: Whether the entities purportedly picked out by past theoretical terms exist or not, there are many respects in which the models specified by past theories are similar to the world. Insofar as past scientists intend their models to represent some target system, the relevant question is not whether past theoretical terms successfully refer but the extent to which past models are similar to their intended target. The semantic view of theories simply eschews the problem of referential failure (e.g., Giere, 1988, p. 107).

Reply: Implicit in Giere’s objection is the claim that a model represents its target system because some scientist intends the model to do so. At least in the case of representational models then, scientists introduce models to represent a target that they have in mind. If a scientist introduces a model to represent the causal origin of a set of observed phenomena, we face the same issues that beset the causal theory of reference. If a scientist introduces a model to represent a target system supposed to satisfy a set of descriptions, we face the same issues that beset the description theory. And if we insist that the establishment of a representational link between a model and its target—or reference-fixing in short—is achieved according to modelers’ intentions, we face the same issues that beset Kitcher's theory. Chakravartty (2007, sec. 7.4) argues that Giere does not escape the general problem of correspondence between a model and its target and that a modeler needs to phrase her epistemic commitments about the model with the help of descriptions. I agree; but I think that the real problem underlying Giere's response is rather that it does not escape the problem of referential failure at all and makes it even less easily tractable.

Objection 2: The context-sensitivity underlying the problem of referential tracking hardly needs a new theory of reference. The specification of a particular context in (C^*) and (T^*) is idle since the relevant context is already fixed by the meaning of the descriptions of the theory. If we select carefully the component descriptions of the core causal-explanatory description associated with a term, they will automatically pick out entities at the appropriate level.

Reply: I must concede that our everyday language is largely insensitive to the type of objective context discussed here. It is implicit in the meaning of 'red' that it applies to macroscopic and perhaps astronomical objects. Otherwise, it is wiser to speak of wavelengths comprised between 620-750nm when we intend to refer to more fine-grained features. It is not true, however, that the same conclusion holds in physics. Taken by itself, Galileo's law of free fall contains no specification about the types of situations to which it applies and does not apply. Similarly, Newton's universal law of gravitation contains no specification to the effect that it does not apply to the gravitational interaction between subatomic entities. The situation is different in physics because at the time a theory is a going concern, the scope of the theory has not yet been delimited by the next theory. We might therefore disagree about the future success of the theory and about the putative referential success of its terms in domains that have not yet been put to the test. A theory of reference that allows us to address the issue of stability must include a condition that specifies the domain where a given theory is likely to remain reliable at a time the theory is still a live option.

Objection 3: The theory of reference (CST^*) does not help selective realists to give a convincing defense of scientific realism. Too many entities that are deemed essential to the explanatory and predictive achievements of theories are specified in ways that extend well-beyond their empirical limits. For instance, the terms 'classical electromagnetic field', 'Higgs field', and 'smoothly curved space-time' appear to refer to entities that are specified by assigning core properties to some target system at arbitrarily short distance scales. (CST*) therefore requires selective realists to be too selective.

Reply: The apparent reference to arbitrarily short distances comes from the putatively fundamental character of the standard formulation of classical electromagnetism, the Standard Model of particle physics, and classical General Relativity. Once we formulate these theories as effective theories, the ambiguity disappears. We can indeed separate the descriptions of these theories according to their probable empirical limitations and still characterize the electromagnetic field, the Higgs field, and the metric field as continuum systems in the appropriate range. Hence, by selecting the continuum descriptions $D(L)$ of these entities for large-distance scales L, we attribute a continuum structure to entities living *at* large-distance scales. Or, to put it differently, the terms 'electromagnetic field', 'Higgs field' and 'smoothly curved space-time' refer to continuous systems *at* large-distance scales. One might think that the continuum structure of these entities enters in conflict with the discrete structure of putatively more fundamental entities. Again, as I argued in section 4.4, this apparent conflict reflects a particular view about ontological reduction, not an issue about referential success. If we take the effective versions of classical electromagnetism, the Standard Model, and General Relativity by themselves, their structure gives us good reasons to believe that 'electromagnetic field', 'Higgs field' and 'smoothly curved space-time' refer to continuous entities at large-distance scales.

4.6 Conclusion

I have argued that the apparent failures of reference over the course of the history of science are best analyzed by examining whether the central theoretical terms of a theory refer to entities specified within the limited physical context where the theory is empirically reliable. Since it is often, if not always, possible to determine at least partially such empirical limitations at the time past theories were still of topical interest, this principle of selective reference enables selective realists to address the challenge that there does not seem to be any principled and reliable way of distinguishing between referential success and failure from the perspective of each theory. I have shown that the framework of effective theories provides us with a paradigmatic set-up for implementing this selective strategy successfully. If we cannot directly use this framework, it still provides us at least with a blueprint for assessing referential success. And in both cases, the suggestion is that instead of implementing a “top-down" strategy by looking at the constituents of a theory that are deemed indispensable to its predictive and explanatory achievements, we should rather implement a “bottom-up" strategy based on the empirical limitations of the theory in order to escape charges of *post hoc* rationalization.

References

Abbott, L. and Wise, M. (1980). "Effective Hamiltonian for Nucleon Decay". *Physical Review D* 22.9, pp. 2208–2212.

Adams, A. et al. (2006). "Causality, Analyticity and an IR Obstruction to UV Completion". *Journal of High Energy Physics* 2006.10, p. 014.

Alonso, R. et al. (2014). "Renormalization Group Evolution of the Standard Model Dimension Six Operators III: Gauge Coupling Dependence and Phenomenology". *Journal of High Energy Physics* 2014.4, p. 159.

Aydemir, U., Anber, M., and Donoghue, J. (2012). "Self-healing of Unitarity in Effective Field Theories and the Onset of New Physics". *Physical Review D* 86.1, p. 014025.

Bagnuls, C. and Bervillier, C. (2001). "Exact Renormalization Group Equations. An Introductory Review". *Physics Reports* 348.1-2, pp. 91–157.

Bain, J. (2000). "Against Particle/Field Duality: Asymptotic Particle States and Interpolating Fields in Interacting QFT (Or: Who's Afraid of Haag's Theorem?)" *Erkenntnis* 53.3, pp. 375–406.

— (2013). "Effective Field Theories". *The Oxford Handbook of Philosophy of Physics*. Ed. by R. Batterman. Oxford: Oxford University Press, pp. 224–254.

Baker, D. J. (2016). "The Philosophy of Quantum Field Theory". *Oxford Handbooks Online*. Philsci preprint: http://philsci-archive.pitt.edu/11375/.

Bangu, S. (2008). "Reifying Mathematics? Prediction and Symmetry Classification". *Studies in History and Philosophy of Science Part B* 39.2, pp. 239–258.

Banks, T. and Zaks, A. (1982). "On the Phase Structure of Vector-like Gauge Theories with Massless Fermions". *Nuclear Physics B* 196.2, pp. 189–204.

Baumann, D. and Green, D. (2012). "Supergravity for Effective Theories". *Journal of High Energy Physics* 2012.3, p. 1.

Baumann, D. and McAllister, L. (2015). *Inflation and String Theory*. Cambridge: Cambridge University Press.

Bertolini, D., Thaler, J., and Thomas, Z. (2013). "Super-Tricks for Superspace". *Searching for New Physics at Small and Large Scales: TASI 2012*. Singapore: World Scientific, pp. 421–496.

Blanchet, L. (2014). “Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries”. *Living Reviews in Relativity* 17.1, p. 2.

Bond, A., Hiller, G., et al. (2017). “Directions for Model Building From Asymptotic Safety”. *Journal of High Energy Physics* 2017.8, p. 4.

Bond, A. and Litim, D. (2017). “Theorems for Asymptotic Safety of Gauge Theories”. *The European Physical Journal C* 77.6, p. 429.

Bond, A., Litim, D., et al. (2018). “UV Conformal Window for Asymptotic Safety”. *Physical Review D* 97.3, p. 036019.

Brivio, I. and Trott, M. (2019). “The Standard Model as an Effective Field Theory”. *Physics Reports* 793, pp. 1–98.

Buchholz, D. (2000). “Algebraic Quantum Field Theory: A Status Report”. arXiv:math-ph/0011044: http://arxiv.org/abs/math-ph/0011044.

Bueno, O. and French, S. (2011). “How Theories Represent”. *The British Journal for the Philosophy of Science* 62.4, pp. 857–894.

Bulycheva, K. and Gorsky, A. (2014). “RG Limit Cycles”. arXiv:1402.2431: https://arxiv.org/abs/1402.2431.

Burgess, C. (2004). “Quantum Gravity in Everyday Life: General Relativity as an Effective Field Theory”. *Living Reviews in Relativity* 7.1, p. 5.

— (2017). “Introduction to Effective Field Theories and Inflation”. arXiv:1711.10592: https://arxiv.org/abs/1711.10592.

Burgess, C. and Moore, G. (2006). *The Standard Model: A Primer*. Cambridge: Cambridge University Press.

Butterfield, J. (2014). “Reduction, Emergence and Renormalization”. *Journal of Philosophy* 111.1, pp. 5–49.

Butterfield, J. and Bouatta, N. (2014). “On Emergence in Gauge Theories at the ’t Hooft Limit”. *European Journal for Philosophy of Science* 5.1, pp. 55–87.

— (2015). “Renormalization for Philosophers”. *Metaphysics in Contemporary Physics*. Ed. by T. Bigaj and C. Wuthrich. Leiden: Brill Rodopi, pp. 437–485.

Calmet, X. and Casadio, R. (2014). “Self-healing of Unitarity in Higgs Inflation”. *Physics Letters B* 734, pp. 17–20.

Campbell-Kelly, M. et al. (2003). *The History of Mathematical Tables: From Sumer to Spreadsheets*. Oxford: Oxford University Press.

Cao, T. Y. and Schweber, S. S. (1993). “The Conceptual Foundations and the Philosophical Aspects of Renormalization Theory”. *Synthese* 97.1, pp. 33–108.

Carroll, S. (2016). *The Big Picture: On the Origins of Life, Meaning, and the Universe Itself*. New York: Dutton.

Castellani, E. (2002). “Reductionism, Emergence, and Effective Field Theories”. *Studies in History and Philosophy of Science Part B* 33.2, pp. 251–267.

— (2019). “Scientific Methodology: A View from Early String Theory”. *Why Trust a Theory?* Ed. by R. Dardashti, R. Dawid, and K. Thébault. Cambridge: Cambridge University Press, pp. 173–183.

Caswell, W. (1974). “Asymptotic Behavior of Non-Abelian Gauge Theories to Two-Loop Order”. *Physical Review Letters* 33.4, pp. 244–246.

Chakravartty, A. (1998). “Semirealism”. *Studies in History and Philosophy of Science Part A* 29.3, pp. 391–408.

— (2007). *A Metaphysics for Scientific Realism: Knowing the Unobservable*. Cambridge: Cambridge University Press.

Chang, H. (2010). “The Hidden History of Phlogiston: How Philosophical Failure Can Generate Historiographical Refinement”. *Hyle* 16.2, pp. 47–79.

Collins, J. (1986). *Renormalization: An Introduction to Renormalization, the Renormalization Group and the Operator-Product Expansion*. Cambridge: Cambridge University Press.

— (2009). “The Problem Of Scales: Renormalization and All That”. arXiv:hep-ph/9510276: https://arxiv.org/abs/hep-ph/9510276.

— (2011). *Foundations of Perturbative QCD*. Cambridge: Cambridge University Press.

Crowther, K. and Linnemann, N. (2017). “Renormalizability, Fundamentality, and a Final Theory: The Role of UV-Completion in the Search for Quantum Gravity”. *The British Journal for the Philosophy of Science* axx052. Philsci preprint: http://philsci-archive.pitt.edu/13219/.

Csaki, C. (2005). “TASI lectures on Extra Dimensions and Branes”. *From Fields to Strings: Circumnavigating Theoretical Physics*. Ed. by M. Shifman, A. Vainshtein, and J. Wheater. Vol. 2. Singapore: World Scientific, pp. 967–1060.

da Costa, N. C. A. and French, S. (2003). *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. Oxford: Oxford University Press.

Darrigol, O. (2000). *Electrodynamics from Ampère to Einstein*. New York: Oxford University Press.

Davey, K. (2003). "Is Mathematical Rigor Necessary in Physics?" *The British Journal for the Philosophy of Science* 54.3, pp. 439–463.

Dawid, R. (2013). *String Theory and the Scientific Method*. Cambridge: Cambridge University Press.

Delamotte, B. (2012). "An Introduction to the Nonperturbative Renormalization Group". *Renormalization Group and Effective Field Theory Approaches to Many-Body Systems*. Ed. by A. Schwenk and J. Polonyi. Springer-Verlag Berlin Heidelberg, pp. 49–132.

Descartes, R. (1991). *The Philosophical Writings of Descartes*. Ed. by J. Cottingham et al. Vol. 3. Cambridge: Cambridge University Press.

Devitt, M. (1984). *Realism and Truth*. 1996, 2nd. Princeton: Princeton University Press.

Donoghue, J. (1995). "Introduction to the Effective Field Theory Description of Gravity". arXiv:gr-qc/9512024: https://arxiv.org/abs/gr-qc/9512024.

— (2020). "Gravitons and Pions". *The European Physical Journal A* 56.3, p. 86.

Donoghue, J., Golowich, E., and Holstein, B. (1994). *Dynamics of the Standard Model*. Cambridge: Cambridge University Press.

Douglas, H. (2009). "Reintroducing Prediction to Explanation". *Philosophy of Science* 76.4, pp. 444–463.

— (2013). "The Value of Cognitive Values". *Philosophy of Science* 80.5, pp. 796–806.

Drake, S. (1964). "Galileo and the Law of Inertia". *American Journal of Physics* 32, pp. 601–608.

Dubovsky, S. et al. (2012). "Effective Field Theory for Hydrodynamics: Thermodynamics, and the Derivative Expansion". *Physical Review D* 85.8.

Duncan, A. (2012). *The Conceptual Framework of Quantum Field Theory*. Oxford: Oxford University Press.

Dymarsky, A. et al. (2015). "On Scale and Conformal Invariance in Four Dimensions". *Journal of High Energy Physics* 2015.10, p. 171.

Earman, J. and Roberts, J. (1999). “Ceteris Paribus: There is No Problem of Provisos”. *Synthese* 118.3, pp. 439–478.

Egg, M., Lam, V., and Oldofredi, A. (2017). “Particles, Cutoffs and Inequivalent Representations”. *Foundations of Physics* 47.3, pp. 453–466.

Eichhorn, A. (2018). “Status of the Asymptotic Safety Paradigm for Quantum Gravity and Matter”. *Foundations of Physics* 48.10, pp. 1407–1429.

— (2019). “An Asymptotically Safe Guide to Quantum Gravity and Matter”. *Frontiers in Astronomy and Space Sciences* 5, p. 47.

Enc, B. (1976). “Reference of Theoretical Terms”. *Nous* 10.3, pp. 261–282.

Endlich, S. et al. (2011). “The Quantum Mechanics of Perfect Fluids”. *Journal of High Energy Physics* 2011.4, p. 102.

Field, H. (1973). “Theory Change and the Indeterminacy of Reference”. *The Journal of Philosophy* 70.14, pp. 462–481.

Forster, M. and Sober, E. (1994). “How to Tell When Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions”. *The British Journal for the Philosophy of Science* 45.1, pp. 1–35.

Franklin, A. (2018). “Whence the Effectiveness of Effective Field Theories?” *The British Journal for the Philosophy of Science* axy050. Philsci preprint: http://philsci-archive.pitt.edu/14731/.

Fraser, D. (2006). *Haag’s Theorem and the Interpretation of Quantum Field Theories with Interactions*. Doctoral Dissertation, University of Pittsburgh. (Unpublished). http://d-scholarship.pitt.edu/id/eprint/8260.

— (2008). “The Fate of ’Particles’ in Quantum Field Theories with Interactions”. *Studies in History and Philosophy of Science Part B* 39.4, pp. 841–859.

— (2009). “Quantum Field Theory: Underdetermination, Inconsistency, and Idealization”. *Philosophy of Science* 76.4, pp. 536–567.

— (2011). “How to Take Particle Physics Seriously: A Further Defence of Axiomatic Quantum Field Theory”. *Studies in History and Philosophy of Science Part B* 42.2, pp. 126–135.

— (2018). “The Development of Renormalization Group Methods for Particle Physics: Formal Analogies Between Classical Statistical Mechanics and Quantum Field Theory”. *Synthese*. Philsci preprint: http://philsci-archive.pitt.edu/14591/.

Fraser, J. D. (2018). “Renormalization and the Formulation of Scientific Realism”. *Philosophy of Science* 85.5, pp. 1164–1175.

— (2020a). “The Real Problem with Perturbative Quantum Field Theory”. *The British Journal for the Philosophy of Science* 71.2, pp. 391–413.

— (2020b). “Towards a Realist View of Quantum Field Theory”. *Scientific Realism and the Quantum*. Ed. by S. French and J. Saatsi. Philsci preprint: http://philsci-archive.pitt.edu/14716/. Oxford: Oxford University Pres.

Frege, G. (1892). “On Sense and Reference”. *Translations from the Philosophical Writings of Gottlob Frege*. Ed. by P. Geach and M. Black. Oxford: Blackwell, 1952.

French, S. (2014). *The Structure of the World: Metaphysics and Representation*. Oxford: Oxford University Press.

Gabbey, A. (1998). “New Doctrines of Motion”. *The Cambridge History of Seventeenth-Century Philosophy*. Ed. by D. Garber. Vol. 2. Cambridge: Cambridge University Press, pp. 649–679.

Galilei, G. (1590). *On Motion, and On Mechanics*. Trans. by I. E. Drabkin and S. Drake. Madison: University of Wisconsin Press, 1960.

— (1632). *Dialogue Concerning the Two Chief World Systems, Ptolemaic and Copernican*. Trans. by S. Drake. Berkeley: University of California Press, 1967.

— (1638). *[Discourses on the] Two New Sciences*. Trans. by S. Drake. Madison: University of Wisconsin Press, 1974.

Gell-Mann, M. (1956). “The Interpretation of the New Particles as Displaced Charge Multiplets”. *Il Nuovo Cimento (1955-1965)* 4.2, pp. 848–866.

Georgi, H. (1992). “Thoughts on Effective Field Theory”. *Nuclear Physics B - Proceedings Supplements* 29.2, pp. 1–10.

— (1993). “Effective Field Theory”. *Annual Review of Nuclear and Particle Science* 43.1, pp. 209–252.

Giere, R. N. (1988). *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

Gies, H. and Jaeckel, J. (2004). “Renormalization Flow of QED”. *Physical Review Letters* 93.11, p. 110405.

Gockeler, M. et al. (1998a). “Is There a Landau Pole Problem in QED?” *Physical Review Letters* 80.19, pp. 4119–4122.

Gockeler, M. et al. (1998b). “Resolution of the Landau Pole Problem in QED”. *Nuclear Physics B - Proceedings Supplements*. Proceedings of the XVth International Symposium on Lattice Field Theory 63.1, pp. 694–696.

Goldberger, W. and Rothstein, I. (2006). “An Effective Field Theory of Gravity for Extended Objects”. *Physical Review D* 73.10, p. 104029.

Gouvea, A. de (2016). “Neutrino Mass Models”. *Annual Review of Nuclear and Particle Science* 66.1, pp. 197–217.

Grandy, R. (1973). “Reference, Meaning, and Belief”. *The Journal of Philosophy* 70.14, pp. 439–452.

Grinbaum, A. (2008). “On the Eve of the LHC: Conceptual Questions in High-energy Physics”. arXiv:0806.4268: http://arxiv.org/abs/0806.4268.

— (2019). “The Effectiveness of Mathematics in Physics of the Unknown”. *Synthese* 196.3, pp. 973–989.

Gripaios, B. (2015). “Lectures on Effective Field Theory”. arXiv:1506.05039: https://arxiv.org/abs/1506.05039.

Gripaios, B. and Sutherland, D. (2015). “Quantum Field Theory of Fluids”. *Physical Review Letters* 114.7, p. 021602.

Gross, D. (1999a). “Renormalization Groups”. *Quantum Fields and Strings: A Course for Mathematicians. Vol. 1*. Ed. by P. Deligne et al. Providence: American Mathematical Society, pp. 551–596.

— (1999b). “The Triumph and Limitation of Quantum Field Theory”. *Conceptual Foundations of Quantum Field Theory*. Ed. by T. Y. Cao. Cambridge: Cambridge University Press, pp. 56–67.

Grzadkowski, B. et al. (2010). “Dimension-six Terms in the Standard Model Lagrangian”. *Journal of High Energy Physics* 2010.10, p. 85.

Haag, R. (1996). *Local Quantum Physics: Fields, Particles, Algebras*. 2nd. Springer.

Halvorson, H. (2007). “Algebraic Quantum Field Theory”. *Handbook of the philosophy of physics, part A*. Ed. by J. Butterfield and J. Earman. North Holland, pp. 731–864.

Hartmann, S. (2001). “Effective Field Theories, Reductionism and Scientific Explanation”. *Studies in History and Philosophy of Science Part B* 32.2, pp. 267–304.

Hellerman, S. et al. (2014). “Effective String Theory Simplified”. *Journal of High Energy Physics* 2014.9, p. 183.

Helling, R. (2012). “How I Learned to Stop Worrying and Love QFT”. Ed. by M. Flory and C. Sluka. arXiv:1201.2714: https://arxiv.org/abs/1201.2714.

Henning, B. et al. (2017). “2, 84, 30, 993, 560, 15456, 11962, 261485, ...: Higher Dimension Operators in the SM EFT”. *Journal of High Energy Physics* 2017.8, p. 16.

Hesse, M. (1961). *Forces and Fields: The Concept of Action at a Distance in the History of Physics*. London: Nelson.

Hollowood, T. (2013). *Renormalization Group and Fixed Points*. SpringerBriefs in Physics. Heidelberg: Springer.

Hooper, W. (1998). “Inertial Problems in Galileo’s Preinertial Framework”. *The Cambridge Companion to Galileo*. Ed. by P. Machamer. Cambridge: Cambridge University Press, pp. 146–174.

Hossenfelder, S. (2013). “Minimal Length Scale Scenarios for Quantum Gravity”. *Living Reviews in Relativity* 16.1, p. 2.

Huggett, N. and Weingard, R. (1995). “The Renormalisation Group and Effective Field Theories”. *Synthese* 102.1, pp. 171–194.

— (1996). “Exposing the Machinery of Infinite Renormalization”. *Philosophy of Science* 63.3, S159–S167.

Jammer, M. (1957). *Concepts of Force: A Study in the Foundations of Dynamics*. Cambridge: Harvard University Press.

Kane, R. (1976). “Nature, Plenitude and Sufficient Reason”. *American Philosophical Quarterly* 13.1, pp. 23–31.

— (1986). “Principles of Reason”. *Erkenntnis* 24.2, pp. 115–136.

Kaplan, J. (2016). “Lectures on AdS/CFT from the Bottom Up”. Available at: http://sites.krieger.jhu.edu/jared-kaplan/files/2016/05/AdSCFTCourseNotesCurrentPublic.pdf.

Keas, M. (2018). “Systematizing the Theoretical Virtues”. *Synthese* 195.6, pp. 2761–2793.

Kitcher, P. (1978). “Theories, Theorists and Theoretical Change”. *The Philosophical Review* 87.4, pp. 519–547.

— (1982). “Implications of Incommensurability”. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1982.2, pp. 689–703.

Kitcher, P. (1993). *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. New York: Oxford University Press.

— (2001). “Real Realism: The Galilean Strategy”. *The Philosophical Review* 110.2, pp. 151–197.

Koyré, A. (1965). *Newtonian Studies*. Cambridge: Harvard University Press.

— (1966). *Galileo Studies*. Trans. by J. Mepham. Atlantic Highlands: Humanities Press, 1978.

Kragh, H. (1990). *Dirac: A Scientific Biography*. Cambridge: Cambridge University Press.

— (2019). “Physics and the Totalitarian Principle”. arXiv:1907.04623: http://arxiv.org/abs/1907.04623.

Kripke, S. (1972). *Naming and Necessity*. Cambridge: Harvard University Press.

Kroon, F. (1985). “Theoretical Terms and the Causal View of Reference”. *Australasian Journal of Philosophy* 63.2, pp. 143–166.

— (1987). “Causal Descriptivism”. *Australasian Journal of Philosophy* 65.1, pp. 1–17.

Kuhlmann, M. (2010). “Why Conceptual Rigour Matters to Philosophy: On the Ontological Significance of Algebraic Quantum Field Theory”. *Foundations of Physics* 40.9-10, pp. 1625–1637.

Kuhn, T. (1977). “Objectivity, Value Judgment, and Theory Choice”. *The Essential Tension*. Chicago: University of Chicago Press, pp. 320–39.

Ladyman, J. (1998). “What is Structural Realism?” *Studies in History and Philosophy of Science Part A* 29.3, pp. 409–424.

— (2011). “Structural Realism versus Standard Scientific Realism: The Case of Phlogiston and Dephlogisticated Air”. *Synthese* 180.2, pp. 87–101.

Ladyman, J. et al. (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.

Landig, A. (2014). “Partial Reference, Scientific Realism and Possible Worlds”. *Studies in History and Philosophy of Science Part A* 47, pp. 1–9.

Laudan, L. (1981). “A Confutation of Convergent Realism”. *Philosophy of Science* 48.1, pp. 19–49.

Lehman, L. (2014). “Extending the Standard Model Effective Field Theory with the Complete Set of Dimension-7 Operators”. *Physical Review D* 90.12, p. 125023.

Lehman, L. and Martin, A. (2016). “Low-derivative Operators of the Standard Model Effective Field Theory via Hilbert Series Methods”. *Journal of High Energy Physics* 2016.2, p. 81.

Lepage, P. (1989). “What is Renormalization?” *From Actions to Answers, Proceedings of the 1989 Theoretical Study Institute in Elementary Particle Physics*. Ed. by T. DeGrand and D. Toussaint. Singapore: World Scientific, pp. 483–509.

Leplin, J. (1975). “The Concept of an Ad Hoc Hypothesis”. *Studies in History and Philosophy of Science Part A* 5.4, pp. 309–345.

Lewis, D. (1984). “Putnam’s Paradox”. *Australasian Journal of Philosophy* 62.3, pp. 221–236.

Litim, D. and Sannino, F. (2014). “Asymptotic Safety Guaranteed”. *Journal of High Energy Physics* 2014.12, p. 178.

Mann, R. B. et al. (2017). “Asymptotically Safe Standard Model via Vectorlike Fermions”. *Physical Review Letters* 119.26, p. 261802.

Manohar, A. (2018). “Introduction to Effective Field Theories”. arXiv:1804.05863: http://arxiv.org/abs/1804.05863.

Martin, S. (2016). “A Supersymmetry Primer”. arXiv:hep-ph/9709356: http://arxiv.org/abs/hep-ph/9709356.

Massimi, M. (2010). “Galileo’s Mathematization of Nature at the Crossroad Between the Empiricist and the Kantian Tradition”. *Perspectives on Science* 18.2, pp. 152–188.

Maxwell, G. (1962). “On the Ontological Status of Theoretical Entities”. *Scientific Explanation, Space, and Time*. Ed. by H. Feigl and G. Maxwell. Vol. 3. Minneapolis: University of Minnesota Press, pp. 3–26.

McLeish, C. (2005). “Scientific Realism Bit by Bit: Part I. Kitcher on Reference”. *Studies in History and Philosophy of Science Part A* 36.4, pp. 668–686.

— (2006). “Scientific Realism Bit by Bit: Part II. Disjunctive Partial Reference”. *Studies in History and Philosophy of Science Part A* 37.2, pp. 171–190.

McMullin, E. (1984). “A Case for Scientific Realism”. *Scientific Realism*. Ed. by J. Leplin. Berkeley: University of California, pp. 8–40.

— (1996). “Epistemic Virtue and Theory Appraisal”. *Realism in the Sciences: Proceedings of the Ernan McMullin Symposium, Leuven, 1995*. Ed. by I. Douven and L. Horsten. Leuven: Leuven University Press, pp. 13–34.

Meheus, J., ed. (2002). *Inconsistency in Science*. Dordrecht: Kluwer Academic Publishers.

Miller, M. (2016). “Mathematical Structure and Empirical Content”. Philsci preprint: http://philsci-archive.pitt.edu/12678/.

Newton-Smith, W. (1981). *The Rationality of Science*. London: Routledge.

Niedermaier, M. and Reuter, M. (2006). “The Asymptotic Safety Scenario in Quantum Gravity”. *Living Rev. Relativity* 9.

Norberg, A. (2003). “Table Making in Astronomy”. *The History of Mathematical Tables: From Sumer to Spreadsheets*. Ed. by M. Campbell-Kelly et al. Oxford: Oxford University Pres, pp. 177–207.

Osborn, H. (2016). “Advanced Quantum Field Theory Lecture Notes”. Available at: http://www.damtp.cam.ac.uk/user/ho/AQFTNotes.pdf.

Perez-Lorenzana, A. (2005). “An Introduction to Extra Dimensions”. *Journal of Physics: Conference Series* 18, pp. 224–269.

Peskin, M. E. and Schroeder, D. V. (1995). *An Introduction to Quantum Field Theory*. Chicago: Westview Press.

Petrov, A. A. and Blechman, A. E. (2016). *Effective Field Theories*. Singapore: World Scientific.

Poisson, E. and Will, C. (2014). *Gravity: Newtonian, Post-Newtonian, Relativistic*. Cambridge: Cambridge University Press.

Polchinski, J. (1988). “Scale and Conformal Invariance in Quantum Field Theory”. *Nuclear Physics B* 303.2, pp. 226–236.

Polchinski, J. and Strominger, A. (1991). “Effective String Theory”. *Physical Review Letters* 67.13, pp. 1681–1684.

Polonyi, J. (2003). “Lectures on the Functional Renormalization Group Method”. *Open Physics* 1.1, pp. 1–71.

Popper, K. (1965). *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Harper & Row.

Porto, R. (2016). “The Effective Field Theorist’s Approach to Gravitational Dynamics”. *Physics Reports* 633, pp. 1–104.

Psillos, S. (1997). “Kitcher on Reference”. *International Studies in the Philosophy of Science* 11.3, pp. 259–272.

— (1999). *Scientific Realism : How Science Tracks Truth*. London: Routledge.

Psillos, S. (2012). “Causal Descriptivism and the Reference of Theoretical Terms”. *Perception, Realism and the Problem of Reference*. Ed. by A. Raftopoulos and P. Machamer. Cambridge: Cambridge University Press, pp. 212–238.

Putnam, H. (1973). “Explanation and Reference”. *Conceptual Change*. Ed. by G. Pearce and P. Maynard. Dordrecht: D. Reidel Pub. Co, pp. 196–214.

— (1975). *Mathematics, Matter, and Method*. Philosophical Papers, Vol. 1. London: Cambridge University Press.

Redhead, M. (1999). “Quantum Field Theory and the Philosopher”. *Conceptual Foundations of Quantum Field Theory*. Ed. by T. Y. Cao. Cambridge: Cambridge University Press, pp. 34–40.

Richter, B. (2006). “Theory in Particle Physics: Theological Speculation versus Practical Knowledge”. *Physics Today* 59.10, pp. 8–9.

Rivat, S. (2019). “Renormalization Scrutinized”. *Studies in History and Philosophy of Science Part B* 68, pp. 23–39.

Rivat, S. and Grinbaum, A. (2020). “Philosophical Foundations of Effective Field Theories”. *The European Physical Journal A* 56.3, p. 90.

Robinson, D. (1992). “Renormalization and the Effective Field Theory Programme”. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992, pp. 393–403.

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* 66, pp. 118–134.

Ruetsche, L. (2011). *Interpreting Quantum Theories: The Art of the Possible*. Oxford: Oxford University Press.

— (2018). “Renormalization Group Realism: The Ascent of Pessimism”. *Philosophy of Science* 85.5, pp. 1176–1189.

— (2020). “Perturbing Realism”. *Scientific Realism and the Quantum*. Ed. by S. French and J. Saatsi. Oxford: Oxford University Pres, p. 293.

Russell, B. (1905). “On Denoting”. *Mind* 14.56, pp. 479–493.

Saatsi, J. (2005). “Reconsidering the Fresnel-Maxwell Theory Shift: How the Realist Can Have Her Cake and EAT It Too”. *Studies in History and Philosophy of Science Part A* 36.3, pp. 509–538.

Saunders, S. (1993). "To What Physics Corresponds". *Correspondence, Invariance and Heuristics: Essays in Honour of Heinz Post*. Ed. by S. French and H. Kamminga. Boston Studies in the Philosophy of Science. Dordrecht: Springer, pp. 295–325.

Schaffner, K. (1972). *Nineteenth-century Aether Theories*. New York: Pergamon Press.

Scharf, G. (1995). *Finite Quantum Electrodynamics: The Causal Approach*. 2nd. Springer-Verlag Berlin Heidelberg.

Schindler, S. (2018). *Theoretical Virtues in Science: Uncovering Reality through Theory*. Cambridge: Cambridge University Press.

Schulte, O. (2008). "The Co-discovery of Conservation Laws and Particle Families". *Studies in History and Philosophy of Science Part B* 39.2, pp. 288–314.

Schwartz, M. (2013). *Quantum Field Theory and the Standard Model*. Cambridge: Cambridge University Press.

Schweber, S. S. (2015). "Hacking the Quantum Revolution: 1925-1975". *The European Physical Journal H* 40.1, pp. 53–149.

Sellars, W. (1963). "Philosophy and the Scientific Image of Man". *Science, Perception and Reality*. 1991st ed. Atascadero, Calif: Ridgeview Publishing Company, pp. 1–40.

Shankar, R. (1999). "Effective Field Theory in Condensed Matter Physics". *Conceptual Foundations of Quantum Field Theory*. Ed. by T. Y. Cao. Cambridge: Cambridge University Press, pp. 47–55.

Siegfried, R. (2002). *From Elements to Atoms: A History of Chemical Composition*. Philadelphia: Amer Philosophical Society.

Smart, J. J. C. (1963). *Philosophy and Scientific Realism*. London: Routledge.

Sober, E. (2008). *Evidence and Evolution: The Logic Behind The Science*. Cambridge: Cambridge University Press.

Stanford, P. K. (2003a). "No Refuge for Realism: Selective Confirmation and the History of Science". *Philosophy of Science* 70.5, pp. 913–925.

— (2003b). "Pyrrhic Victories for Scientific Realism". *Journal of Philosophy* 100.11, pp. 553–572.

— (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford: Oxford University Press.

Stanford, P. K. (2015). ““Atoms exist" is Probably True, and Other Facts That Should Not Comfort Scientific Realists”. *Journal of Philosophy* 112.8, pp. 397–416.

Stanford, P. K. and Kitcher, P. (2000). “Refining the Causal Theory of Reference for Natural Kind Terms”. *Philosophical Studies* 97.1, pp. 97–127.

Steinmann, O. (2000). *Perturbative Quantum Electrodynamics and Axiomatic Field Theory*. Springer-Verlag Berlin Heidelberg.

Sterelny, K. (1983). “Natural Kind Terms”. *Pacific Philosophical Quarterly* 64.2, pp. 110–125.

Streater, R. F. and Wightman, A. S. (2000). *PCT, Spin and Statistics, and All That*. Princeton: Princeton University Press.

Strocchi, F. (2013). *An Introduction to Non-Perturbative Foundations of Quantum Field Theory*. International Series of Monographs on Physics. Oxford: Oxford University Press.

Summers, S. J. (2016). “A Perspective on Constructive Quantum Field Theory”. arXiv:1203.3991: https://arxiv.org/abs/1203.3991.

Sundrum, R. (2005). “TASI lectures: To the Fifth Dimension and Back”. arXiv:hep-th/0508134: http://arxiv.org/abs/hep-th/0508134.

Teller, P. (1988). “Three Problems of Renormalization”. *Philosophical Foundations of Quantum Field Theory*. Ed. by H. R. Brown and R. Harre. Oxford University Press, pp. 73–89.

— (1989). “Infinite Renormalization”. *Philosophy of Science* 56.2, pp. 238–257.

— (1995). *An Interpretive Introduction to Quantum Field Theory*. Princeton: Princeton University Press.

Thagard, P. R. (1978). “The Best Explanation: Criteria for Theory Choice”. *The Journal of Philosophy* 75.2, pp. 76–92.

United States Naval Observatory. Nautical Almanac Office (1855). *American Ephemeris and Nautical Almanac*.

van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: Clarendon Press.

Vickers, P. (2013). *Understanding Inconsistent Science*. Oxford: Oxford University Pres.

Wallace, D. (2006). “In Defence of Naivete: The Conceptual Status of Lagrangian Quantum Field Theory”. *Synthese* 151.1, pp. 33–80.

Wallace, D. (2011). “Taking Particle Physics Seriously: A Critique of the Algebraic Approach to Quantum Field Theory”. *Studies in History and Philosophy of Science Part B* 42.2, pp. 116–125.

— (2019). “The Quantum Theory of Fields”. *The Routledge Companion to the Philosophy of Physics*. Ed. by E. Knox and A. Wilson. Forthcoming. Philsci preprint: http://philsci-archive.pitt.edu/15296/. Routledge.

Weinberg, S. (1979a). “Baryon- and Lepton-Nonconserving Processes”. *Physical Review Letters* 43.21, pp. 1566–1570.

— (1979b). “Ultraviolet Divergences in Quantum Theories of Gravitation”. *General Relativity: An Einstein centenary survey*. Ed. by S. W. Hawking and W. Israel. Cambridge: Cambridge University Press, pp. 790–831.

— (1995). *The Quantum Theory of Fields, Vol. 1: Foundations*. Cambridge: Cambridge University Press.

— (1996). *The Quantum Theory of Fields, Vol. 2: Modern Applications*. Cambridge: Cambridge University Press.

— (1999). “What is Quantum Field Theory and What Did We Think It Was”. *Conceptual Foundations of Quantum Field Theory*. Ed. by T. Y. Cao. Cambridge: Cambridge University Press, pp. 241–251.

— (2000). *The Quantum Theory of Fields, Vol. 3: Supersymmetry*. Cambridge: Cambridge University Press.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

Wells, J. D. (2012a). “Effective Field Theories and the Role of Consistency in Theory Choice”. arXiv:1211.0634: http://arxiv.org/abs/1211.0634.

— (2012b). *Effective Theories in Physics: From Planetary Orbits to Elementary Particle Masses*. Heidelberg: Springer Science & Business Media.

— (2018). “Lexicon of Theory Qualities”. *Physics Resource Manuscripts*. Available at: https://sites.lsa.umich.edu/jwells/manuscripts/.

Westfall, R. S. (1971). *Force in Newton’s Physics: The Science of Dynamics in the Seventeenth Century*. New York: American Elsevier.

White, J. H. (1932). *The History of the Phlogiston Theory*. London: E. Arnold & co.

Williams, P. (2015). “Naturalness, the Autonomy of Scales, and the 125 GeV Higgs”. *Studies in History and Philosophy of Science Part B* 51, pp. 82–96.

— (2019a). “Renormalization Group Methods”. *The Routledge Companion to the Philosophy of Physics*. Ed. by E. Knox and A. Wilson. Forthcoming. Philsci preprint: http://philsci-archive.pitt.edu/15346/. Routledge.

— (2019b). “Scientific Realism Made Effective”. *The British Journal for the Philosophy of Science* 70.1, pp. 209–237.

Wilson, K. G. (1971). “Renormalization Group and Strong Interactions”. *Physical Review D* 3.8, pp. 1818–1846.

Worrall, J. (1989). “Structural Realism: The Best of Both Worlds?” *Dialectica* 43.1-2, pp. 99–124.

Zinn-Justin, J. (2007). *Phase Transitions and Renormalization Group*. Oxford: Oxford University Press.