

When do we stop digging?

Conditions on a fundamental theory of physics

Karen Crowther
Department of Philosophy
University of Geneva

Abstract

In seeking an answer to the question of what it means for a theory to be *fundamental*, it is enlightening to ask why the current best theories of physics are *not* generally believed to be fundamental. This reveals a set of conditions that a theory of physics must satisfy in order to be considered fundamental. Physics aspires to describe ever deeper levels of reality, which may be without end. Ultimately, at any stage we may not be able to tell whether we've reached rock bottom, or even if there is a base level—nevertheless, I draft a checklist to help us identify when to stop digging, in the case where we may have reached a candidate for a final theory. Given that the list is—according to (current) mainstream belief in high-energy physics—complete, and each criterion well-motivated, I argue that a physical theory that satisfies all the criteria can be assumed to be fundamental in the absence of evidence to the contrary.

1 Introduction

It may come as a surprise, but not one of our current theories of physics are generally considered fundamental.¹ This is in spite of their being among the most accurate and successful scientific theories ever. It is in spite of our best theory of matter—the standard model of particle physics—allegedly describing all the fundamental particles and forces! And it is in spite of our best theory of spacetime—general relativity—describing what are surely the most fundamental features of our universe imaginable: space and time. Physicists are not content to rest with these theories; they expect more-fundamental physics to lie beyond—to be buried deeper underneath.

In this essay, I consider why physicists are apparently so hard to satisfy—that is, I ask why these two theories are not thought to be fundamental. The aspects of these theories that are responsible for their reputation as non-fundamental can then be inverted, and framed as constraints that any theory must satisfy if it *is* to be considered fundamental. Thus, I compile a checklist of necessary conditions on a fundamental theory, according to common belief in high-energy physics. I argue that the conditions in the checklist—although they are

¹While quantum chromodynamics is often cited as a fundamental theory, I argue that this belief is incompatible with the full list of conditions on a fundamental theory.

not all able to be precisely defined, nor rigorously justified—reflect the essential character of physics itself. Thus, given that the list is—from the perspective of current high-energy physics—complete, and each criterion well-motivated, I argue that a theory of physics that satisfies all the criteria can be assumed to be fundamental in the absence of evidence to the contrary.

The structure of the essay is as follows: In §2, I outline some of the different ideas of *fundamentality* associated with modern physics, before, in §3, explaining why neither the standard model of particle physics nor general relativity (GR) are typically considered fundamental. Following this, §4 presents the necessary conditions that physicists apparently place on a fundamental theory and explores what these reveal about the nature of physics, as well as its dream of a final theory. In §5, I consider the implications of these conditions for *quantum gravity* (QG)—the (currently unknown) theory that is supposed to be more fundamental than GR and quantum field theory. And, conversely, I also investigate what QG reveals about our conception of fundamentality.

2 Fundamental and more-fundamental in modern physics

Conceptions of fundamentality in modern physics have been heavily shaped by the framework of *effective field theory* (EFT), and its associated philosophy. The framework of EFT is a means of constructing theories—called *effective field theories*—that are each valid only at a given “level”, i.e., at large distances (corresponding to low energy scales, since energy is inversely proportional to length) compared to a particular short length (high-energy) scale “cutoff”, Λ . To do this, we can start either with a short-distance theory and then use the toolbox of EFT to produce a theory that describes the same system at larger distances, or the other way around (from a large-distance theory to one that describes the system at shorter-length scales), or we can start from scratch to construct a theory at the required scale when no shorter- or longer-distance one is available. Effective theories are generally not considered to be fundamental, because they break down (i.e., cease to be predictive) at lengths approaching Λ . The picture presented by this framework is of a “tower of theories”, each valid at a different level. Each theory is framed in terms of the appropriate parameters for its particular level, representing the physical interactions that are important at that level, and these parameters are specific to each theory—they do not exist at the levels above or below in the tower [1].

Because of this, there is a compelling case to be made that what is fundamental depends on the level we are interested in—for each level there is a theory that clearly describes the relevant physics of the system being studied, and is framed in terms of the appropriate parameters [2]. While physicists generally believe that we could *in principle* use a shorter length scale theory in order to make predictions about the system at some particular larger distance scale (i.e., if we had access to the required computational resources, plus the ability to use them), to do so would not only be very (and needlessly) complicated, but would also hopelessly obscure the picture of the relevant physics at the scale of interest. We do not need to use a theory of atoms in order to describe a game of baseball, for instance. And yet, there is an asymmetry between the levels—we believe that the laws we use to describe the motion of the baseball after being hit *depend on* those of atomic physics, but not that the laws of

atomic physics depend on the laws used to describe the baseball. Atomic theory is thought to be *more fundamental* than the laws we use to predict the trajectory of a baseball—the atoms and their interactions are “more basic” than any macroscopic objects that they compose.

This asymmetry is captured in the way we move between theories: It is generally believed that *in principle*, with full knowledge of the physics of a system at a particular short-length scale (plus, again, the required computational resources and the ability to use them), we could arrive at results valid at any larger scales without requiring any additional information. On the other hand, the large-scale physics *underdetermines* the shorter-scale theory: We could not, *even in principle*, derive the correct theory of a system at small-length scales from a complete description of its physics at some larger length scale. More information would be required. For this reason, the tower of theories is usually thought to be ordered *hierarchically*, with the shorter-scale theories being more fundamental, and so lower on the tower, than the larger-scale, “higher-level” ones. I use this notion of *relative fundamentality* here.²

We encounter problems, however, when we attempt to move from relative to absolute terms—what does it mean for a theory to be *fundamental* rather than just more- or less-fundamental? How do we define “rock bottom” of the tower? We might suppose it would be a theory valid at the tiniest length scales. Yet, according to quantum field theory (QFT)—the framework within which the standard model of particle physics is formulated (and, as I explain below, is also understood in terms of EFT)—there is an arbitrarily large amount of energy available in the vacuum, and so the tower may be endless, “shorter and shorter-length turtles all the way down”, thus implying that there is no final theory [3].³ Commonly, it is thought that the way to escape this conclusion is to recognise, though, that we are not trapped in the framework of QFT [4, 5, 6]. It is possible that we find a theory that is not a QFT⁴, and which yields predictions at the tiniest length scales.

Although stepping outside the framework of QFT may free us from worrying about an arbitrarily large amount of energy available in the vacuum,⁵ however, it does not save us from the possibility that there is still new physics beyond any theory that we arrive at. The recognition that we are not trapped in the framework of any given theory produces an *epistemic worry*: Even if we reach a theory that yields predictions for the tiniest length scales (or, equivalently, all possible high-energy scales), we cannot be sure that it actually describes the world at these scales (unless, of course, we have access to experimental data at all possible high energy scales!)

A theory (whether a QFT or not) that is formally predictive at all possible high energy scales is said to be *UV complete* (“UV” referring to the short-wavelength end of the electromagnetic spectrum, the ultraviolet). Although I take it that a theory being UV complete

² As we shall see, however, equating “more fundamental” with “shorter-distance”, is problematic because the very idea of *distance* may cease to be applicable at some point, and yet we may have reasons to expect there to be another, presumably more fundamental, theory *beyond* that point—i.e., beyond the domain of space and time—which could then not be called a “shorter-distance” theory!

³Energy and length scales are inversely related, so an arbitrarily large energy scales correspond to arbitrarily short length scales.

⁴Indeed, as I discuss below, QFT is not considered a fundamental framework, and so it is expected that a fundamental theory will not be a QFT.

⁵This arbitrarily large vacuum energy may, in fact, be interpreted as an artifact of a non-fundamental formalism (§3.1).

is *necessary* for its being fundamental,⁶ being UV complete is not *sufficient* for a theory to be fundamental—it does not guarantee that there is “nothing beyond” [7]. For example, consider Newton’s laws of motion: These are formally predictive in all domains, and yet we know that they are not *correct* at all scales. Another example is quantum chromodynamics (QCD, the theory of the strong nuclear force), which is UV complete and yet, as I discuss below, is not considered fundamental.⁷

So, a theory formally being predictive to all high-energy scales, and thus apparently being the lowest brick in the tower (or, at least, one of the bricks at the lowest level of the tower), is no guarantee that it is in fact a fundamental theory—UV completeness alone is not enough reason to stop digging. Yet, it is one constraint on a fundamental theory.⁸ In order to understand what kind of theory would motivate physicists to stop digging, and to answer the question of what it means for a theory to be fundamental, I now invert it: Why do we not consider our current best theories of physics to be the final word? Why are we currently digging for a more fundamental theory?

3 Why our current best theories of physics are not fundamental

3.1 Quantum field theory and the standard model

We must distinguish between the *framework* of QFT, and *particular QFTs* (i.e., the theories formulated within this framework). To begin with, consider the framework of QFT; there are three inter-related reasons it is not considered fundamental. Firstly, the framework is mathematically ill-defined, which means, secondly, that, traditionally,⁹ the theories it generates are plagued by singularities—infinities pop up all over the place (e.g., in the arbitrarily large vacuum energy mentioned above). Some of these need to be dealt with in order to render the afflicted theories usable, and the procedure by which this is done is known as *renormalisation*. The third reason the framework is not considered fundamental is that, although it utilises special relativity, it is not *generally relativistic*—it does not take into account our best theory of spacetime. Many physicists interpret this as the cause of the ill-definedness of the framework, as well as the singularities that appear within its theories. The modern interpretation of QFT holds that these singularities are artifacts of a non-fundamental framework, and that their appearance in our QFTs is the result of our ignorance of the more-fundamental physics at shorter-length scales beyond—physics that includes a quantum theory of gravity (discussed below). Thus, QFT is treated as an *effective* framework; i.e., as EFT.

⁶If this were not true, then it would mean that the world is just not amenable to scientific description at extremely high-energy scales. But my arguments in §4 justify our dismissal of this “UV silence scenario” [7].

⁷Physicists usually distinguish between a *fundamental* theory and a *final* theory, arguing that although QCD is not a final theory, its UV completeness means that it is a fundamental theory. I argue below §4 that this reasoning is not consistent with the rest of the conditions on a fundamental theory.

⁸Neglecting the possibility of the UV silence scenario, Footnote 6. Also, I take a “theory without distance” (as in Footnote 2) to be UV complete, in the sense that it does not break down at any short distance scale.

⁹As we shall see, this problem was solved by considering QFT as EFT.

Apart from being products of a non-fundamental framework, there are four reasons why QFTs are considered non-fundamental. (1) The first regards the necessity of renormalisation (for those theories where it is necessary)—the idea is, again, that the singularities dealt with by this procedure are due to our neglect of some shorter-scale (more fundamental) physics beyond. (2) Some theories are believed to be non-fundamental because they are not UV-complete. Such theories, like quantum electrodynamics (QED, the quantum field theory of electromagnetism, describing light and the interactions of charged particles) break down at some short-distance scale, and are thus EFTs.¹⁰ (3) Many QFTs are supposedly non-fundamental because they are not exactly solvable (this is due to the ill-definedness of the framework), and so employ the approximation techniques of *perturbation theory*. (4) Finally, the standard model of particle physics itself is *non-unified*. Although the standard model can be written as a single theory, it appears as a disjointed amalgam of separate (particle) fields, which drives many physicists to seek a more unified theory beyond [8].

This means that QFTs such as QCD that do not suffer any of the difficulties (1–3) at short-distances are still regarded as non-fundamental—Not just because they are products of a non-fundamental framework, but because they are not part of a unified theory.

3.2 General relativity

The reasons for not believing GR fundamental are the motivations for seeking *quantum gravity* (QG)—the as-yet-undiscovered theory needed to describe physics in the domains where both QFT and GR are thought to be necessary. These domains include, for instance, the *Planck scale*, which is the unfathomably small distance of 10^{-32} cm. The theory is expected to replace GR, and describe the more-fundamental physics that “underlies” spacetime. Since QG is supposed to be a quantum theory of spacetime (i.e., a theory that takes into account both quantum theory and GR) one of its motivations is the desire for unification, as well as a desire for a *unique* theory, rather than multiple frameworks. Another factor driving the search for QG is the presence of (particular types of) singularities in GR such as black hole singularities, and the “big bang” singularity. These are “places” where the theory is formally (mathematically) ill-defined—apparently representing a breakdown of spacetime—and QG is supposed to shed light on these.

4 Conditions on a fundamental theory

According to the above discussion, a fundamental theory must be:

- Unified;
- Unique;
- UV complete (“nothing beyond” formally);
- Non-perturbative (exactly solvable);
- Internally consistent (well-defined formally, with no problematic singularities)

¹⁰In the case of QED, this is due to the presence of a Landau pole divergence.

Several of these ideas reflect a more general principle: *That a fundamental theory not leave anything apparently in need of explanation.* For instance, if a theory is not UV complete, or otherwise not well-defined everywhere, then we are led to ask what happens in the domains that the theory does not describe. More generally, if a theory is not internally consistent, or relies on approximations, physicists tend to believe that this is a symptom of there being something missing—some physics that the theory fails to take into account.

Consider if we did not have a *unique* theory, but a “patchwork” of several (UV-complete, and otherwise apparently fundamental) theories. These would have to fit together in an especially particular way, such that there were neither any gaps, nor any overlap in the domains of the world covered by these theories. Otherwise, if there were gaps, we would ask about the phenomena not described by any of the theories, or the physics “between” the theories—we would search for a description of this. And, if there were overlap, with a particular phenomenon described by more than one theory, then we would ask which (if any) provides the more fundamental description. Thus, if the patchwork of several apparently fundamental theories did not match up perfectly, we would be led to search for a more fundamental theory.

However, if the patchwork *did* match up perfectly—such that, at the smallest distance scales, all physical phenomena were covered, and there was only one description of each phenomenon—we could “stitch” these all together to form a single theory.¹¹ In this case, the theory would be unique, and satisfy all conditions for fundamentality, *except for unification.* The idea of unification is not just that there be a single theory describing all phenomena, but that it describe all phenomena as *the same*—as fundamentally stemming from a single origin, e.g., as manifestations of a single entity or interaction.

The requirement of unification is hard to justify. Given that our manifest experience of the world is of diversity rather than a sameness of phenomena, seeking an explanation of heterogeneity seems counter-intuitive—surely a *unified* description would be more striking than a disunified one, and cry out for explanation? I return to discuss this below. For now, though, I add a few more criteria to the list that are also motivated by the need for explanation. Generally—according to high-energy physics—it is believed that a fundamental theory must also be:

- Level comprehensive (“no gaps and no overlap” in description at the scales that the theory is required¹² in order to describe);
- Background independent (no fixed structures across all models of the theory);
- Natural (no “fine-tuning” of parameters);
- Not weird

The first of these additional criteria—which I call *level comprehensiveness*—stems from the need for a complete and non-overlapping¹³ description of the physics at the most fundamental level. Full coverage of *all physics* is then supposed to be ensured by this condition,

¹¹Thus, I argue that a fundamental theory should be unique, *contra* the typical distinction drawn between a fundamental and final theory, according to which only the latter need be unique (Footnote 7).

¹²Note that there may be overlap in lower-energy, less-fundamental descriptions.

¹³Descriptive redundancies such as gauge symmetries are acceptable; the existence of dualities is a more complicated issue, however [9].

plus the criterion of UV completeness *and* the belief (or assumption), discussed above, that results about any low-energy phenomena can *in principle* be obtained from the higher-energy theory.

The notion of *background independence* has several aspects (including that the theory be non-perturbative), but the general idea is that there are no *fixed* (“background”) structures in the theory—nothing that has to be specified *for* the theory “by hand” instead of being determined *by* the theory itself [10, 11]. Otherwise, the appearance of such structures requires explanation, and implores us to seek a more fundamental theory that provides this.

Naturalness is, basically, the requirement that a theory not contain dimensionless numbers that are either very large or very small; it means that a large-distance theory does not depend sensitively on the choice of parameters in the more-fundamental (higher energy-scale) theory [12, 13]. Else, these parameters will appear to have been fine-tuned: Standing out—to the physicists’ eye—as “unnatural” and in need of explanation.

Finally, the last requirement, *no weirdness*, is the more general idea (underlying the criteria of background independence and naturalness) that the theory not say anything that causes sufficient unease amongst physicists to drive the community to seek an explanation for the “weirdness”. (An example is the *measurement problem* in quantum theory, which is disquieting enough that many researchers believe that a more-fundamental theory must solve it [14]).

All nine conditions above assume a lot about the world—for instance, why should it be everywhere amenable to physical description, and why should this description be within our ability to formulate? Why should the world be such that our theories of it are formally neat and mathematically tractable, rather than messy and unusable? Why should these theories be “reasonable” to us, rather than strange and disquieting? And, as I asked above, why does manifest diversity and disunity, rather than covert underlying unity, require explanation? Each criterion needs further justification, especially if its standardly-cited motivation is an imagined trajectory of physics based on a particular reading of the history of physics—as, for instance, tending towards greater unification [8], or, as Weinberg [15] believes, a “convergence of the arrows of explanation”.

These are all good philosophical questions for which I have no answers. What I argue, however, is that each condition—while it may not be precisely definable nor rigorously philosophically justifiable—nevertheless captures something central to the enterprise of physics itself. Physics *does* and *must*, by its nature, assume that we are able to formulate a physical description of all phenomena, and that this description is useful to us as far as it can be. If physics were to abandon this assumption, then it would seem to have “given up”, in a sense: We would no longer be doing physics. It is also key to the “business of physics” that it explain diverse phenomena by appeal to simple, universal laws. It’s just what physics does. And this underlies its requirement that a fundamental theory be unified.

The list of conditions has been compiled from the perspective of physics itself: It represents the criteria that any theory must satisfy if it is to be counted, *by physics*, as a fundamental theory. Thus, I argue that, given that this list is complete, a physical theory’s fulfilment of these conditions is sufficient for that theory to be treated as fundamental according to physics. Note, however, that this list was drawn up based on current theories, and it is possible that these views change; for instance, QG may very well force us to reconsider our requirements.

Before turning to QG, however, I must emphasise that the above conditions are just those that a theory of physics—i.e., a *scientific* theory—must satisfy in order to be considered fundamental. In other words, I am presupposing that any candidate fundamental theory of physics already satisfies some further conditions such that it is acceptable as a scientific theory. I do not consider these further conditions (which *may* include requirements relating to predictions, experiment, falsifiability, etc.)—suffice to say they are even more controversial, especially given the current state of QG research.

5 Quantum gravity

As stated in §3.2, the motivations for QG include the desire for unification, as well as the desire for a unique theory. Yet, QG it is not necessarily a *unified* theory, nor a *theory of everything*. In the first case, it may be a *semiclassical* theory, which is a non-unified combination of GR and QFT, and in the second case, it may just be a quantum theory of gravity, and not a theory that combines gravity with the standard model forces. In other words, although it is expected to be more-fundamental than GR and QFT, *QG is not necessarily a fundamental theory* [7].

Currently, there is no QG. Instead, there are a number of *approaches* to finding a theory (i.e., different research programs), of various stages of development, with none, as yet, yielding novel, experimentally testable predictions. One of these approaches, *string theory*, alleges to be a unified theory of everything. An additional claim for string theory’s being a candidate for a fundamental (i.e., final) theory is its UV completeness—using a symmetry in the theory known as “T-duality”, an argument can be made that string theory describes a minimal length [16]. Thus, string theory apparently says there can be no new physics “beyond” that described by the approach. According to string theory, then, we should stop digging.

But, as described above (§1), even if string theory is correct, its UV completeness does not alleviate the epistemic worry: It is possible that new physics exists beyond, and that a more-fundamental theory be necessary (one not featuring the minimal length described by string theory). In other words, it is possible that string theory, even if correct, turns out to be an *effective* theory after all.

String theorists, however, believe that this is not the case, and that the approach will produce a final theory.¹⁴ The problem, though, is that the approach does not satisfy many of the other conditions on a fundamental theory—in particular, it is only known *perturbatively*, it is background *dependent*, and its parameters are not uniquely determined (string theory is not a single theory, but a huge “landscape” of possible theories). Additionally, the approach is not level comprehensive, because of overlaps—different theories (known as *dual* theories) can potentially be interpreted as describing the “same physics” [9]. All of these factors drive string theorists to search for a more-fundamental type of string theory underlying the currently-known “versions” of it. (Researchers in other approaches to QG, however, are duly sceptical of a positive outcome).

¹⁴Arguments for this appear in [16].

6 Conclusion

I have drafted a list of nine criteria that a fundamental theory of physics must satisfy, according to physics itself. I argued that a physical theory's fulfilment of these necessary conditions is jointly *sufficient* for it to be regarded, by current physics, as fundamental, while still recognising that there may be additional conditions discovered in the future. The epistemic worry that nags us to always keep digging for a more-fundamental theory—the worry that, no matter what theory we arrive at, there might still be new physics beyond—can be turned around, thanks to the two general principles that underlie the nine conditions. These principles are: Full, non-overlapping coverage of description, and comprehensiveness of explanation. If we believe these are satisfied, then the question shifts from “What if there's something beyond?” to “Why should we think there is something beyond?” That is, the burden of justification is transferred. If a theory is found that satisfies all these conditions, and yet its status as fundamental is disputed, then we can ask why. The answer will either be unacceptable, or it will reveal further conditions to add to the list. Currently, however, we are still digging.

References

- [1] Howard Georgi. Effective-field theory. *Annual Review of Nuclear and Particle Science*, 43:209–252, 1993.
- [2] Philip W. Anderson. More is different. *Science*, 177:393–396, 1972.
- [3] Tian Yu Cao and Silvan S. Schweber. The conceptual foundations and the philosophical aspects of renormalization theory. *Synthese*, 97(1):33–108, 1993.
- [4] Elena Castellani. Reductionism, emergence, and effective field theories. *Studies In History and Philosophy of Modern Physics*, 33(2):251–267, 2002.
- [5] Karen Crowther. *Effective Spacetime: Understanding Emergence in Effective Field Theory and Quantum Gravity*. Springer, Heidelberg, 2016.
- [6] Nick Huggett and Robert Weingard. The renormalisation group and effective field theories. *Synthese*, 102(1):171–194, 1995.
- [7] Karen Crowther and Niels Linnemann. Renormalizability, fundamentality and a final theory: The role of UV completion in the search for quantum gravity. *British Journal for the Philosophy of Science*, 2017.
- [8] Tim Maudlin. On the unification of physics. *Journal of Philosophy*, 93(3):129–144, 1996.
- [9] Elena Castellani and Dean Rickles, editors. Special Issue on Dualities in Physics. *Studies in History and Philosophy of Modern Physics*, 59:1–142, 2017.
- [10] Jeremy Butterfield and Christopher Isham. On the emergence of time in quantum gravity. In J. Butterfield, editor, *The Arguments of Time*, pages 116–168. Oxford University Press, Oxford, 1999.
- [11] Lee Smolin. The case for background independence. In D. Rickles, S. French, and J. Saatsi, editors, *The Structural Foundations of Quantum Gravity*, pages 196–239. Oxford University Press, Oxford, 2006.
- [12] Gian F. Giudice. Naturally speaking: The naturalness criterion and physics at the lhc. In G. Kane and A. Pierce, editors, *Perspectives on LHC Physics*. World Scientific, 2008.
- [13] James D. Wells. The utility of naturalness, and how its application to quantum electrodynamics envisages the standard model and Higgs boson. *Studies in History and Philosophy of Modern Physics*, 49:102–108, 2015.
- [14] Wayne Myrvold. Philosophical issues in quantum theory. *The Stanford Encyclopedia of Philosophy*, (<https://plato.stanford.edu/archives/spr2017/entries/qt-issues/>), 2017.
- [15] Steven Weinberg. Newtonianism, reductionism and the art of congressional testimony. *Nature*, (330):433–437, 1987.
- [16] Richard Dawid. *String Theory and the Scientific Method*. Cambridge University Press, 2013.