Defining a crisis: The roles of principles in the search for a theory of quantum gravity*

Karen Crowther
University of Geneva

Abstract

In times of crisis, when current theories are revealed as inadequate to task, and new physics is thought to be required—physics turns to re-evaluate its principles, and to seek new ones. This paper explores the various types, and roles of principles that feature in the problem of quantum gravity as a current crisis in physics. I illustrate the diversity of the principles being appealed to, and show that principles serve in a variety of roles in all stages of the crisis, including in motivating the need for a new theory, and defining what this theory should be like. In particular, I consider: the generalised correspondence principle, UV-completion, background independence, and the holographic principle. I also explore how the current crisis fits with Friedman's view on the roles of principles in revolutionary theory-change, finding that while many key aspects of this view are not represented in quantum gravity (at the current stage), the view could potentially offer a useful diagnostic, and prescriptive strategy. This paper is intended to be relatively non-technical, and to bring some of the philosophical issues from the search for quantum gravity to a more general philosophical audience interested in the roles of principles in scientific theory-change.

Keywords: Constitutive principles; Scientific revolution; Theory change; Friedman; Holographic principle; Spacetime

Contents

---

1 Introduction

In times of crisis—when current theories are revealed as inadequate to task, and new physics is thought to be required—physics turns to re-evaluate its principles, and to seek new ones. This was recognised by both Poincaré and Einstein in their work leading up to, and during, the crises in physics at the turn of the 20th century that resulted, eventually, in the development of relativity and quantum mechanics (among numerous other innovations). Fast forward to today, and physics is again—or, perhaps, *still*—in strife. The crisis it faces now is that of *quantum gravity*, and, in a way, it is a continuation of the older crises.[1] The problem of quantum gravity is the quest for a theory (QG) needed to describe the domains in which both general relativistic and quantum field theoretic effects are supposed to play a nontrivial role. These domains include, for instance, those where curvatures and energy densities approach the Planck scale, (10^{-35}m), e.g., near black hole singularities, and cosmological singularities (such as the "big bang"). Part of the difficulty with finding a theory is the extreme nature of these regimes, which preclude direct experimental testing (although tests in accessible regimes are not ruled out), and currently there are no unequivocal data that QG is definitely required to explain. This empirical disconnect means that more weight has fallen on other guides to theory construction and evaluation, including principles. Another great difficulty, not unrelated to the first, is that it is unclear what an acceptable theory of QG would look like. Apart from the minimal characterisation offered above, there is little agreement as to the criteria of theory success. Accordingly, there are many different approaches to QG (i.e., different research programs), which are distinguished, in part, by the differing sets of principles that they adopt.[2]

In 1984, John Wheeler wrote that, "[n]o question about quantum gravity is more difficult than the question, 'What is the question?'" (Wheeler, 1984, p. 224)—and, while much progress has been made in the decades since, the most difficult question remains. In order to answer this question, we need to consider the various principles

---

[1] Almost as soon as GR was completed, Einstein was aware of the need for a quantum theory of gravity; Einstein (1916, p. 202) writes of a possible conflict between GR and the principles of quantum theory. See also the papers in Blum & Rickles (Forthcoming).

[2] Note that, although each of these approaches is incomplete, and faces its own set of problems in addition to these general difficulties, it is still possible—and indeed, worthwhile—to engage with the philosophy of QG, as well as to consider what QG might tell us about other philosophical questions (See, e.g., Butterfield & Isham, 2001; Callender & Huggett, 2001; Rickles, 2008a).

that motivate the search for the new theory, as well those that serve to define it. Developing and evaluating these principles is particularly important, given the long-standing theoretical stalemate.[3] The aim of this paper is to explore the variety of different types of principles that feature in the search for the new theory, as well as their different roles and statuses. I work with only a minimal characterisation of "principle", as being any conceptual or formal feature of physical theorising that is taken to play a key role in the definition, construction and/or justification of new theories.[4] This general definition is based on how the term is used in physics itself, and is necessary in order to accommodate the variety of putative principles that currently feature in the search for the new theory.

Notably, this definition admits many more instances than the specific notions standardly discussed in the philosophy literature do—for instance, it includes items like unification, mathematical consistency, and the generalised correspondence principle, which are apparently difficult candidates for first principles, conventional definitions, or constitutive principles, but instead seem more like meta-principles, or virtues.[5] However, at this stage of theory development, I maintain that it is important to consider all such putative principles initially—and, after examining them, we will be in a better position to evaluate whether or not they accord with any particular philosophical interpretation or view of scientific principles. It is only once we are clear on what, and how, principles are being used in this context that we will be able to engage in both the descriptive and prescriptive tasks, and do so from both directions—exploring what we can learn about, and for, particular philosophical positions given the current "live" case-study in physics, and what we can learn about, and for, the future development of physics, given particular philosophical views. In this paper, I can only make a start on this much larger project. Thus, here I begin the task of exploring some of the key principles that have been appealed to in the search for QG, and I do so *in a neutral way*, without being tied to any particular philosophical view or interpretation of principles in science.

Additionally, though, I also begin to consider how the current crisis in physics fits with Michael Friedman's view of physical theories (as in, e.g., Friedman, 2001)—not only because this is a very influential recent account of the roles of principles in theory-change, but because this view potentially offers an interesting perspective on the problem of QG, as being the *incoherency of physics' current framework*, taken as a whole (in Friedman's sense of "framework", defined below, §2.2). I find that some aspects of what Friedman calls *communicative rationality*—which is a particular process of argument (engaged in by the scientists) that facilitates theory-change—

---

[3]See also, Crowther & Rickles (2014); Crowther & Linnemann (2017); Smolin (2017).

[4]When I refer to "current principles", I mean those that feature within, or led to the development of, existing theories (typically the current best theories of physics). This is in contrast with *new principles* that are being appealed to (e.g., the holographic principle), which have not yet led to, or featured within, any successful, accepted scientific theories.

[5]As I discuss, the generalised correspondence principle does feature as a sort of meta-principle in what Friedman (2001) calls "communicative rationality".

do feature clearly in the context of QG research. However, many other aspects of Friedman's views are not currently reflected in the present crisis, in spite of suggestions that it will necessitate revolutionary paradigm change (as explained in §3). I consider the utility, too, of translating Friedman's thesis into a prescriptive strategy for use in the search for a new physical framework.

The paper begins (§2) with an exploration of the various different types, and roles, of principles in physical theorising—particularly in theory construction and justification. A subsection (§2.1) also mentions Einstein's distinction between principle theories versus constructive theories, as well as Poincaré's views on physical principles as conventions, and his 1904 lecture describing the then-present crisis of physics—I will later (§3) contrast these views with the current crisis in physics, and approaches to theory-development. A following subsection (§2.2), briefly outlines the relevant aspects of Friedman's views that I consider in the rest of the paper. I then (§3) present the problem of QG, explaining the various roles played by different principles in motivating, and defining, the new theory, and how these serve (in part) to characterise the different approaches to QG. Following this (§4) I discuss some of the key principles involved in QG: the correspondence principle (§4.1), UV-completion (§4.2), background independence (§4.3), and the holographic principle (§4.4). As I explain, each of these is of different origin and status. The utility and consequences of any particular principle depends on the way it is employed in the search for the new theory, as well as the other principles that are adopted with it.

2 Principles in Physical Theorising

In physics, there are some putative principles that are understood as applying only within a restricted domain: perhaps as features of effective theories[6] or frameworks, as "effective principles" (such as the principle of least action), or being the result of particular boundary conditions (such as the second law of thermodynamics). There are others that are embodied only within specific theories, and are difficult to imagine featuring in other theories, or being in any way "prior" to theory (such as Pauli's exclusion principle). And, still, there are some principles that are thought to be fundamental—and thus (strictly speaking) to hold in all domains—yet are not useful in all domains, nor even practically applicable in most (e.g., Heisenberg's uncertainty principle, or fundamental symmetries that are broken, via, e.g., the Higgs mechanism). While these are still principles according to the definition given above (i.e., they are conceptual or formal features of physical theorising that play key roles in theory definition, construction and justification)—typically when I refer to "principles", I mean overarching statements that transcend particular theories, and are of broad practical applicability. These principles, such as the principle of relativity, and the correspondence principle, boast a degree of generality and a privileged status that means that we do not expect

---

[6]An *effective theory* in physics is one that is valid only at a given range of energy (length) scales, and is thus not considered fundamental.

any decent theory of physics to be in contradiction with them. Rather, they are used to help arrive at viable physical theories.

The ways in which principles are used in the development of a theory are various, and may serve to confer, or reflect, their different statuses—notwithstanding the fact that a principle may be used in different ways in different research programs, and at different stages of theory-development within a given program. There are at least four different roles of principles in theory development and evaluation:

1. A *guiding principle* is one used primarily as a heuristic: it can aid in the discovery of the theory by leading to new insights, but may or may not actually be retained in the resulting theory (e.g., Mach's principle in the development of GR);

2. A *postulate* is explicitly, and from the outset, taken to be a pillar of the theory (e.g., the speed of light in special relativity);

3. A *criterion of theory acceptance* (also referred to as *criterion of theory success*, or a *constraint*): the new theory should not be accepted if it is incompatible with the principle (unless there is strong evidence in favour of the theory, and/or the principle is shown to be violated under the relevant conditions). Such constraints may be taken either,

    i. as part of the definition of the theory (some popular candidates in different approaches to QG include, e.g., UV-completeness;[7] general covariance; unification), or,

    ii. as "external" constraints (perhaps coming from current successful theories, in their known domains of success; e.g., a theory of QG should not violate Lorentz invariance in any detectable way);

4. A principle may also be appealed to as a *means of confirmation*: compatibility with the principle serving to increase credence in the theory in a way broadly similar to empirical data. This compatibility may be,

    i. *expected*, e.g., if the principle also features in any of the other roles 1–3 (e.g., general covariance in loop quantum gravity; compatibility with Lorentz symmetry in causal set theory), or,

    ii. *unexpected*, as a sort of "bonus", or something like a "prediction" of the new theory (e.g., if a theory of QG were to *automatically* recover, or explain, particular principles of current physics, e.g., the equivalence principle; Lorentz symmetry; the quantum-mechanical uncertainty relations; etc.;

---

[7]Although UV-completion (the idea that a theory be formally predictive up to all possible high energy scales) is often presented as part of the definition of QG, Crowther & Linnemann (2017) argue against taking it as a criterion of acceptance in QG. Nevertheless, it plays a role in motivating the search for QG, and may usefully act as a guiding principle.

but, importantly, this role is not restricted to current principles, as discussed below—unexpected compatibility with *new* principles can also serve to increase credence in the theory).

As should be clear, the roles listed are certainly not exclusive, nor the distinctions between them sharp. This will be more apparent in the discussion of the three examples below (§4).

A little further discussion of role 4. is required. The idea is that principles can serve as means of *non-empirical confirmation*. However, the role 4. is *not the same* as the means of, and arguments for, non-empirical confirmation presented by Dawid (2013). There is a similarity, nevertheless, between role 4.ii. and Dawid's (2013) *unexpected explanatory coherence argument* (UEA), in that the unexpected recovery, or explanation of a particular known principle (i.e., one from current successful theories) would satisfy both Dawid's UEA as well as my role 4.ii. The main differences between these ideas, though, are that, firstly, Dawid's UEA is more specifically about the new theory drawing *explanatory connections* between features of current theories, rather than just recovering, or being compatible with, particular principles of current physics.[8] Secondly, unlike Dawid's UEA, my role 4.ii. can be served also by *new* principles— not just those recognised in current physics. I discuss the idea of new principles more below. But first, I briefly explain Dawid's (2013) *meta-inductive argument* (MIA).

The MIA holds that credence in a new theory is increased by the theory's employment of particular features of our current best theories—we know that these current theories are successful, and so we have reason to believe that a new theory that also uses these known features and techniques, is itself likely to be successful. Thus, a new theory employing a known principle from successful current physics would have its credence increased according to both the MIA and my role 4.i. Though, again, these two ideas are not the same (in particular, role 4.i. can be fulfilled by new principles). Dawid (2013) shows how the MIA plays a role in the non-empirical confirmation of string theory.[9] The standard model of particle physics is an extremely successful model of quantum field theory (QFT), and string theory can be viewed as an extension of this research program. Empirical support for the standard model, Dawid (2013) argues, can be viewed as support for string theory, because string theory is (in a sense) a continuation of the QFT framework, and employs the principles of this framework. (This is important for my arguments in §3).

Returning to the list above: we might wonder whether a given principle that is appealed to as a means of confirmation (4.) must not also already be taken as a constraint (3.)—for otherwise, if it is not a principle that the theory *must* satisfy, how could it be presented as evidence for the theory? Dawid's (2013) MIA shows one way that this can be the case: past success of a principle, used in the context of a different

---

[8]Note, too, that Dawid (2013) presents the UEA in conjunction with two other arguments, in order to establish limits on the *underdetermination* of the new theory, and together paint a picture that is much more nuanced than what I present here.

[9]But see Footnote 8.

theory, can increase credence in a new theory that also features this principle. But past success alone is not enough for us to adopt a principle as a constraint (3.) on the new theory—especially in the case of QG, given the extreme domains where the theory is supposed to apply, as well as the general belief (induced by various heuristic arguments, and supported by the challenging results of many attempts to construct a theory) that QG will represent a *revolutionary* transition. We cannot assume—or, indeed, require—that the new theory conform to current principles[10] (except to say that the new theory not conflict with these principles in a detectable way where the current principles are known to be successful). We need well-motivated, independent reasons for adopting a particular principle in role 3.

Nevertheless, a new theory featuring a known principle from successful theories would increase our credence in that theory, and make us more likely to accept it (4.i.). An example of this is perhaps unification—it is not part of the definition of QG that it be a unified theory of everything (as mentioned above, §1, and below, §3, the defining constraint is just that the theory cover the domains of necessity of GR and quantum theory), yet if a theory turned out to also do this, then we would judge it favourably— partly because of the belief that unification featured in past successful theories (as well as other, independent arguments for approving of more unified theories). Thus, although it may seem a sensible condition on a principle that if it is to feature in role 4. it must already feature in role 3., this may be too restrictive, in that it excludes unforeseen principles that may turn out to be desirable and serve to increase our credence in the theory.

Thus, we might be prompted to include an additional role:

5. A *fallible constraint* is a non-necessary (i.e., not featuring in role 3.) desideratum; able to be appealed to as a general aspiration (or a guiding principle, role 1.), as well as a means of confirmation (4.); it represents a positive feature that we would be happy if our theory possessed, and thus serves to increase our credence in the theory, yet it is not a feature that is *required* of the theory, nor necessarily even a *known* principle.

2.1 Poincaré and Einstein on principles

In this subsection, I briefly describe the views of Poincaré and Einstein in regards to the roles of physical principles in times of crisis. Later (§3), I consider how these ideas might apply in the context of the current crisis in physics.

In September 1904, Poincaré delivered an address to the International Congress of Arts and Science (World's Fair) in St. Louis, Missouri, in which he described two crises in physics.[11] The first crisis—already passed—had led theoretical physicists to

---

[10]Dawid (2013), and string theorists, of course, would challenge this statement, asserting string theory's success in doing exactly this. However, string theory is currently not at a stage where such assertions represent compelling arguments.

[11]English translation, Poincaré (1905a); also appears as Chapters VII–IX of Poincaré (1907).

turn away from studying the detailed structure of the objects, forces, and mechanisms responsible for the phenomena, and instead to consider the general principles of physics, since these, Poincaré says, are capable of yielding truths about the phenomena without requiring knowledge of such details. As a result of this crisis, a new physics had thus emerged: "the physics of the principles". Poincaré (1905a, p. 5) lists the six general principles that formed the basis of this new physics, and states that their application to different phenomena is sufficient for us to learn of the phenomena "all that we could reasonably hope to know of them". In modern terms, these principles are: conservation of energy, the second law of thermodynamics, Newton's third law, relativity, conservation of mass, and the principle of least action.

For Poincaré, these principles possess a generality, certainty and permanence because they are *conventions*: adopted freely (though not arbitrarily) as definitions. Although such principles initially come from laws constructed to describe the results of experiments or observations, once they are adopted as conventions, they are placed beyond the realm of confirmation and disconfirmation—no experiment can ever contradict them.[12] And yet, in 1904, Poincaré recognises that this new physics is itself in peril—this is the second crisis he describes, and it seems to threaten each of the principles, save the sixth. He considers each of the principles in turn, and explains how they are apparently being challenged by recent observations of various phenomena, as well as the results of particular experiments.[13] As Poincaré (1905a) notes, the principles being (even apparently) threatened by experiment seems to contradict his view of their status as conventions. Poincaré's response to an anticipated criticism on this point has a few different lines—the main one being a re-emphasis of his view that a principle can always be shielded from refutation through the addition of *ad hoc* auxiliarly assumptions. But, Poincaré insists that such measures to save a principle are unacceptable in science because they render the principle useless—and, Poincaré (1905a, p. 22) states that, "if a principle ceases to be fecund, experiment without contradicting it directly will nevertheless have condemned it". While this response may not actually answer the criticism posed, it demonstrates that there are limitations to Poincaré's views on the status of conventions: if a principle is so insulated from the empirical realm that it cannot lead to any knowledge, then it is worthless. Conventions, then, should only be retained so long as they are useful. If our principles are indeed broken, then we must not seek to mend them in this way; instead, Poincaré (1905a) says, it would be necessary to begin anew.

Einstein (1919) distinguishes between two types of physical theories. Most theories, he says, are *constructive theories*, which build up a description of complex phenomena from basic results (constituents). Such theories produce an *understanding* of the phenomena, and are comprehensive, adaptable and clear. *Principle theories*, on the other hand, are are not built from constituents, but by appeal to empirically observed

---

[12]See, e.g., Poincaré (1905b), Ch. VI; Poincaré (1907), Ch. X. 4.

[13]Including Brownian motion, radioactivity, the Michelson-Morley experiment, and experiments involving electrons accelerated to very high velocities.

general properties of phenomena; the merits of these theories are their logical perfection, and the security of their foundation. Relativity (both the special, and general theories) is an example of a principle theory. Although Einstein (1919) does not view principles as conventions, his statements here are in line with some of the ideas expressed by Poincaré—in particular, the idea that principles are elevated from empirical observations, such that we believe no experiment could contradict them, and the idea that principles may be appealed to when the details are obscure. While it is certainly the case in QG that the constituents are unknown and inaccessible, remarkably— as Smolin (2017) notes—most attempts at QG are constructive theories. Approaches such as string theory, loop quantum gravity (LQG), non-commutative geometry, causal set theory (and other "discrete approaches"), for instance, all can be seen as aiming to "build up" gravitational phenomena from hypothetical constituents. Nevertheless, principles are still being used in the various roles listed above. Perhaps, then, one of the challenges of QG is to advance a principle theory—one that begins with the postulation of novel principles, in the way that special relativity and GR did. One way of understanding this may be for the theory to postulate a novel *constitutive principle*, as suggested by Friedman (in, e.g., Friedman, 2001).

2.2 Friedman's thesis

A more recent view on principles in science, that has been popular and influential, is that of Friedman, who develops a Neo-Kantian view of constitutive principles, inspired by Poincaré and Reichenbach. According to this view, scientific knowledge has three levels (pp. 45–6 Friedman, 2001). At the base level are the empirical parts of physical theories—the concepts, principles, and laws that are directly applied and tested. The second level are the constitutive principles. These are special, Friedman says, in that they define a space of conceptual possibilities, and so determine the framework of investigation that enables us to devise and express the empirical parts of the theories. The constitutive principles come in two varieties: *mathematical principles*[14], which define the mathematical framework that allows certain kinds of physical theories to be constructed, and *coordinating principles*[15], which mediate between the abstract mathematical structures and the concrete physical phenomena to which the theories are supposed to apply. The framework defined by the constitutive principles is what Kuhn (1962) calls a *paradigm*: the "at least relatively stable sets of rules of the game, as it were, that define or make possible the problem solving activities of normal science including, in particular, the rigorous formulation and testing of properly empirical laws" (Friedman, 2001, p. 45).

A scientific revolution that comes as a result of crisis will correspond, Friedman says, to a change in the constitutive principles. This is effected by the work that occurs at the next level of science: the third level, which comprises "philosophical meta-paradigms"

---

[14]Cf. Samaroo (2015).

[15]An idea attributed to Reichenbach and Schlick, see Friedman (2001, pp. 76–9).

that motivate and sustain the transition from one framework to another. Thus, Samaroo (2015) refers to *Friedman's thesis* as the claim that revolutionary theory change proceeds by deliberate philosophical reflection on constitutive principles. Friedman develops the example of GR: at the empirical level are the Einstein field equations, which are possible thanks to the mathematical constitutive principle of Riemannian manifolds, plus the coordinating constitutive principle, which Friedman argues is the equivalence principle (i.e., that it serves to bridge the Einstein field equations with the empirical realm). Constitutive principles are of the same origins as Poincaré's conventions: elevated from laws based on empirical observations. As such, any new constitutive principle—to define a new framework and make possible a revolution—has its origins in the current (i.e., predecessor) framework. The adoption of the new principle, and the corresponding transition from the old to new frameworks is a gradual process that is implemented by the exercise of *communicative rationality*. This is a means of argument that aims at achieving agreement, and proceeds by appealing to patterns of reasoning that are acceptable to all parties in the dispute (Friedman, 2001, pp. 54–5).

Communicative rationality is employed in two directions: facilitating and legitimating the move to the *new* paradigm, as well as looking back and showing that the *old* paradigm (and the theories within it) was also rational (the idea being to demonstrate that the progress of science is, on the whole, a rational process). *Prospective* communicative rationality may be achieved by appealing to the fact that the new constitutive principle is rooted in a known (i.e., generally accepted) empirical result (Friedman, 2001, pp. 101–3). Additionally, part of prospective communicative rationality involves framing the new paradigm as a "natural continuation" of the (concepts and principles of the) preceding one. Part of the activity that occurs at Friedman's third level of science during a revolutionary transition is the philosophical reflection on what this idea of "natural continuation" means (Friedman, 2001, pp. 105–7). As I discuss below (§3), this is indeed occurring in the case of QG—however, it is happening in such a fractured way that it is perhaps not properly classed as communicative rationality at all. *Retrospective* communicative rationality is achieved by demonstrating that the old paradigm (or, really, theories within it) is contained within the new one, as an approximate special case; as (Friedman, 2001, p. 96) says, "[i]ndeed, with respect to the purely mathematical part of our constitutive frameworks, we have the stronger result that the later principles contain the earlier principles quite exactly, in that the space of possibilities in mathematics continuously (and, as it were, monotonically) increases." This relates to the generalised correspondence principle (§4.1), which is being used in QG research in many different ways—and yet, it is *not* being appealed to in order to demonstrate the rationality of the theories that QG is seeking to replace. Thus, QG is not—at least currently—engaging in retrospective communicative rationality, and, indeed, it would seem strange if it were.

In fact, as I show, this idea of the new theory (i.e., the theory-in-development) "containing" the older one(s) is being used instead in roles 1.–4. above, in the construction and justification of the new theory—i.e., it features heavily as part of *prospective* com-

municative rationality. Some other aspects of Friedman's thesis also do not fit so well with the current situation in physics. For instance, most of the principles being used in the search are not constitutive—although some could plausibly become constitutive principles, or lead the way to new constitutive principles. These latter possibilities, however, also conflict with Friedman's (2001) view on the origin of constitutive principles, which are supposed to be observations or laws elevated from the empirical realm—a possibility that is currently dim in the case of QG. Additionally, the most likely candidate for a new constitutive principle—the holographic principle (§4.4)—does not have any empirical support. Interestingly, too, the most prominent approach towards QG—string theory (§3)—arguably does not utilise any new constitutive principle, and thus may not represent a new framework (and hence nor a revolutionary transition) in Friedman's sense (although it is plausibly a new framework in the more familiar sense in physics).

3   Quantum Gravity

Unlike Poincaré's "second crisis", the present search for a new physical theory is not driven by unusual observations or inexplicable experimental results given our current physics. Indeed, both GR and QFT are incredibly well-confirmed experimentally—no test has ever conflicted with either of them. Because of this, the reasons for seeking a new theory stem rather from conceptual discontent than (practical) physical necessity.[16] These motivations include, for instance, the desire for *unification*—for physics to provide a consistent picture of the world at the fundamental level.[17] There is also the pull of intellectual curiosity, and the ambition to understand ever more of the universe—even if this means theoretically probing regions that will forever remain inaccessible to us physically. As stated above, there are domains that both GR and QFT are expected to be necessary in order to describe, and so we need a new theory that adequately captures the relevant insights of both. Additionally, within each of these frameworks are suggestions that neither is the final word—including, e.g., the quantum-mechanical measurement problem, (particular) GR singularities, and some conceptual dissatisfaction with the framework of QFT related to mathematical inconsistency and the necessity (in some theories) of renormalisation. A very quick (and crude) explanation of this last point: perturbative QFT is infamously plagued by infinities, which are dealt with by the procedures of regularisation and renormalisation. The modern interpretation of QFT understands these theories as *effective field theories*, valid only at relatively large distance scales, and the infinities as a result of our ignorance of the physics at small-length scales beyond. To many physicists' minds, QG, if it provides a minimal length scale (i.e., a finite cut-off for length), could provide some concrete physical justification of these procedures.

---

[16]See, e.g., Crowther (2016, §1); Rickles (2008a, §3).

[17]Although, as Crowther & Linnemann (2017) point out, QG need not be a fundamental theory, it would represent progress in this direction.

The need for a new theory in order to satiate our intellectual curiosity, however, is not itself the crisis. And neither do any of the perceived deficiencies of current theories amount to a crisis. Instead, the crisis emerges from the attempt to construct the new theory, which reveals the apparent necessity of a scientific revolution. The indications for this come from the sheer difficulty of constructing a satisfactory theory, the failure of all familiar methods to do so, and the strange results that come from such attempts— all of which can be seen as symptoms of a fundamental conflict of principles between GR and QFT. The results most often cited as suggestive of the need for a radically new framework are the heuristic arguments for a minimal length (Hossenfelder, 2013).

One of the arguments for a minimal length comes from combining the characteristic constants of the two theories. A particle of mass m has its Compton wavelength, $l_C = \hbar/mc$, equal to its Schwarzschild radius, $l_S = Gm/c^2$, at the Planck mass, $m_{Pl} = \sqrt{\hbar c/G}$, where G is Newton's gravitational constant, $\hbar$ is the reduced Planck's constant, and c is the speed of light. A particle's Compton wavelength is a prediction of QFT, which states that localising m to within l_C uses enough energy to create another (identical) particle of mass m, which results in an indeterminacy in the number of particles present. The Schwarzschild radius is a prediction of GR: it states that compressing m to within the distance l_S will result in the formation of a black hole.Thus, probing the Planck scale, m_{Pl}, is expected to lead to an indeterminate number of microscopic black holes. In addition, these micro black holes are expected to be rapidly-evaporating, based on Hawking's (1974) calculations using QFT in curved spacetime (another means of combining aspects of GR and QFT in order to approximate the predictions of QG).

Considering a quantum description of the gravitational field, Wheeler spoke of a "quantum foam" at the Planck scale, where quantum fluctuations of spacetime (or fluctuations affecting spacetime) would become significant—geometry at this scale is thought to be ill defined, or "fuzzy" (Wheeler & Ford, 1998). The idea that spacetime could cease to be a meaningful concept beyond some scale is a profound one, that departs radically from the current pictures presented by both GR and QFT (as well as manifest experience), and conflicts with almost every principle that features in these frameworks (insofar as they rely on spacetime, space, and/or time).

But, the two frameworks themselves present contrary accounts of spacetime.[18] In GR, spacetime geometry is obtained by solving the Einstein field equations for the metric. Because the metric is a dynamical variable that is also responsible for spacetime geometry (i.e., the kinematic structure against which physical processes are defined), it follows that geometry itself is dynamical: we have to solve the dynamics in order to get to the kinematics. This is related to the theory's alleged background independence (§4.3). On the other hand, QFT appears to be necessarily background dependent in a more substantive way than GR, relying on a fixed spacetime geometry, in the sense that it is the same for all models of the theory (i.e., it does not vary from solution to solution depending on how the matter-energy content of the universe looks). These

---

[18]For more detail see, Butterfield & Isham (1999, 2001); Rickles (2008a).

12

conflicting accounts of spacetime have traditionally been thought to lead to a number of deep puzzles that appear to crop up in many contexts within QG research, under monikers such as the *problem of time* and the *problem of change*.[19] (Note, however, several aspects of these problems arise even at the classical level in GR, and here they may in fact be exposed as nonissues.[20])

There are other deep conflicts between the principles of the two frameworks[21], for instance the principle of *unitarity*, which is central to quantum theory, is not meaningful in GR on the standard way of understanding the theory. This principle has two aspects: that every initial quantum state evolves to a unique final state (i.e., unitary evolution takes pure states to pure states), and (the consequence) that the sum of probabilities of all possible outcomes of any quantum event equals unity at any given time, i.e., a fixed time in a preferred slicing of spacetime. In GR, however, there are no preferred foliations (ways of slicing spacetime). So, the concept of unitarity can only be introduced by specifying a foliation, using the *Hamiltonian formulation* of the theory: if the initial value problem is well defined in this formulation, then the spacetime is said to be globally hyperbolic, and the theory is, in a sense, unitary. If it is not, then, the theory is non-unitary.

The principle of unitarity is frequently cast as a gladiator in the ring by various paradoxes resulting from (apparently reasonable) combinations of GR and QFT—for instance, the *information loss paradox*[22] (one form of which turns on a violation of global hyperbolicity coming from the specific means of combining GR and QFT), and the *firewall paradox*[23]. In the latter, unitarity is pitted against the equivalence principle of GR. Because the gladiators are each favoured by different sections of the crowd, the victor, it is thought, will only be decided by QG—the emperor who is yet to be identified. Yet, I suggest that the way of finding QG could potentially be through the selection of gladiators, which may be achieved by careful analysis of the motivations for favouring each one, and the consequences of keeping them alive.

Meanwhile, there are some principles that have been thought to be put in peril by the combination of GR and QFT, on account of their differing (or being thought to differ) in the two frameworks—and yet may not actually be threatened in this way. These include, for instance, *causality* and *locality*. In QFT, causality is axiomatic (spacelike separated bosonic fields must commute, and spacelike separated fermionic fields must anti-commute), while in GR, it is determined by the matter distribution. Combining these theories suggests that light cones (determining the causal structure of spacetime) will fluctuate with spacetime. On the other hand, however, it seems prima facie possible that, on another way of combining these frameworks, the fixed causal structure of QFT could provide an upper bound on the dynamical causal structure of

---

[19]See, e.g., Anderson (2012); Huggett et al. (2013); Rickles (2006a).

[20] See, e.g., Pitts (2014, 2017); Pons et al. (2010); Thbault (2012).

[21]I take some of these examples from (Rickles, 2008a, §3.6.1), see also, Butterfield & Isham (2001).

[22]See Mathur (2009) for a technical review; and Belot et al. (1999); Wallace (2018) for introductions more accessible to philosophers.

[23]Almheiri et al. (2013).

GR.[24]

In regards to locality: in QFT, observables are localised in spacetime,[25] while it is often thought (i.e., on some standard, but not universal, ways of interpreting gauge) that observables in GR are non-local, due to the theory's diffeomorphism invariance. Put very basically, (one aspect of) the idea is that diffeomorphism invariance prevents one from localising observables to particular points in spacetime (and other field values in the vicinity), because there are equivalent ways of doing so, to different points of spacetime—i.e., location is apparently not a gauge-invariant quantity, and to construct gauge-invariant observables in GR, one has to take into consideration the whole space-time. While several authors have variously argued that this is problematic, others have apparently diagnosed the source of the confusions, and (claim to) have shown (in different ways) how local observables in GR may be defined.[26] Thus, it seems this particular issue may not carry-over to QG, and locality is not necessarily threatened by a conflict of principles (although the problem of defining local observables in QG also arises, in a different manner, if the theory does not feature spacetime (Huggett & Wüthrich, 2013)).

The disparities of principles (whether actual or merely apparent) are all part of the crisis of QG—which itself can be seen as the problem of finding the correct set of principles that the new theory is to satisfy. Thus, we can view the crisis from the relatively immediate perspective of the researchers involved—as Poincaré did at the start of the 20th century—where framing the problem is primarily about settling the constraints (role 3.) on the new theory (a process that starts by reflecting upon each of the current principles that is caught in the fray). Actually finding a theory, though— being the ultimate solution to the crisis—involves also choosing the right postulates (2.) and guiding principles (1.). And for the new theory to be accepted, we additionally need principles serving in roles 4. and 5. (although these are likely to be various, and unable to be identified prior to their use). As should be clear, the principles whose conflicts are taken to signal the presence of the crisis are not just constitutive principles (arguably, this was also the case for the crises described by Poincaré (1905a)). And, as I show below (§4), many of the principles being used in the search for the new theory (in roles 1.–5.) are not constitutive, either—however, they could plausibly become, or lead the way to, constitutive principles.

Nevertheless, it is also interesting to consider the current situation in physics from the perspective of Friedman's view—where the crisis can be understood as *an inco-herent framework* (in Friedman's sense) of modern physics when *taken as a whole*, i.e., the frameworks underlying GR, QFT, and perhaps also statistical physics, are not consistent with one another. This is a problem that can usually be ignored, since (the theories within each of) these frameworks each cover different domains of phenomena,

---

[24]Thanks to a referee for pointing this out.

[25]However, physically relevant QFTs also contain non-observable fields, some of which are non-local.

[26]Highlighting the problems are, e.g., Belot & Earman (2001); Earman (2006); Pooley (2015); Rickles (2006b, 2008b, 2012). Arguments for local observables are in, e.g., Bergmann & Komar (1960); Pitts (2014, 2017); Pons et al. (2010); Pons (2005); Rovelli (1991, 2002).

and mostly we work within one domain at a time. But the crisis of QG is provoked by exactly seeking to describe physics in the domains where all three frameworks are (apparently) required. The consequences of the inconsistency, manifest in these domains, are then interpreted as suggesting the necessity of a revolution—an entirely new framework.

And yet, such a revolution may not actually be necessary. There are a number of possible ways of constructing a theory of QG that do not require an entirely new framework (in Friedman's sense), and are thus non-revolutionary. One way that is frequently discussed is to develop a *semiclassical* theory: a hybrid combination of quantum matter and classical gravity. Such a pairing may not be elegant, but—arguably—may not be be impossibly inconsistent, either.[27] Another approach to QG that would not represent a new framework is the *asymptotic safety scenario*, which shows how GR treated in the framework of QFT may yield a viable (i.e., predictive) theory that describes physics at the Planck scale.[28] Additionally, a theory of QG that builds heavily upon one or the other of the two existing frameworks may not necessarily end up representing (or requiring) a revolution. Such a theory deals with the inconsistency and conflicts of principles in the current framework by simply rejecting some current principles, rather than positing new ones.

Arguably, string theory is an example of this. As mentioned above (§2), string theorists take a key selling-point of their theory to be its status as a natural extension of QFT—the framework that produced the incredibly successful standard model of particle physics. The program aims to avoid the problems (e.g., the infinities mentioned above) inherent to QFT, related (in part) to the locality of the theory, by positing 1-dimensional extended fundamental objects, "strings", rather than pointlike interactions on a background spacetime (i.e., one fixed in the same sense as in QFT, as described above). This spacetime differs from that described by GR, not just by being fixed, however, but by also having many more dimensions. String theory is about the paths that strings sweep out as they move through spacetime. These paths are 2-dimensional surfaces, called *world sheets*. So, the novel insight that characterises string theory (and could thus potentially represent its constitutive principle) is the move from 1-dimensional world lines in QFT, to 2-dimensional world sheets in string theory.

In fact, we can easily switch between the QFT and string theory descriptions of the physics. To see this, consider that mathematically, a world sheet is a 2-dimensional surface embedded in some N-dimensional spacetime. The physical position of each point on the world sheet is given by N numbers specifying the coordinates of the point, and these N quantities form N fields on the world sheet. Now, instead of thinking of a

---

[27]See, e.g., Huggett & Callender (2001); Mattingly (2005, 2006, 2009); Wüthrich (2005).

[28]GR treated in the framework of QFT is apparently *perturbatively* non-renormalisable, breaking down at energies approaching the Planck scale, and so cannot represent a theory of QG, if QG is required to describe physics at this scale. The idea of asymptotic safety is that this apparent problem with GR treated as a QFT may be an artifact of the misapplication of perturbation theory, and that the theory is in fact non-perturbatively renormalisable. Reviews: Niedermaier & Reuter (2006); Percacci (2009).

string as a 2-dimensional object moving in an N-dimensional spacetime, we can think of it as N fields on the 2-dimensional world sheet—a system that is described by QFT. In other words, string theory in an N-dimensional spacetime can be viewed equivalently as a QFT of N interacting fields on a 2-dimensional spacetime.[29]

Note that this alternative formulation should not obscure the fact that string theory is still significantly conceptually novel compared to QFT. Arguably, however, the theory does not depart radically enough to represent a new framework in Friedman's sense: its constitutive basis is the same as that of QFT. The failure of string theory to properly take into account the insights of GR at this stage—in particular, the idea of background independence—is one of the main criticisms levelled at the program from those in opposing approaches, for instance, LQG. Yet, there are some ways in which string theory can be seen as moving towards a background independent approach (as discussed below, §4.3). If this process eventually results in the formation of a new paradigm, then it would align with Friedman's views.[30] Actually, there is perhaps a stronger case to be made, that very few of the current approaches to QG would count as a new framework in Friedman's sense. But, we might instead think to propose Friedman's thesis as a prescriptive strategy in the case of QG—encouraging the development of new constitutive principles (e.g., the holographic principle, §4.4), rather than just the sacrifice of old ones, and promoting communicative rationality between the different approaches.

As it stands, there are some aspects of communicative rationality in play—for instance, most approaches to QG do claim to proceed conservatively, and present their theory as a natural extension of current ones (recall that this is also central to Dawid's MIA). Nevertheless, radical results, especially those that accord with the existing (theoretical and heuristic) indications of physics changing dramatically at the Planck scale, tend to be promoted as "predictions", or means of justifying the theory (i.e., role 4.ii., even if these results are not empirically testable).[31] But this is not actually communicative rationality, because it does not appeal to methods acceptable to all parties involved—instead, string theory, for instance, sells itself as a natural extension of QFT, while LQG is promoted[32] as starting from GR and taking familiar steps to produce a quantum theory. In order to move forward, we need, as Friedman would say, to shift the discussions to the third level of science, and attempt to come to an understanding of what counts as a "natural extension" (I argue that this is part of seeking a more precise definition of the *generalised correspondence principle*, which features heavily in all roles of theory construction and justification in QG, §4.1).

In order to evaluate any of the potential theories of QG *as potential theories of QG*

---

[29]Actually, string theory is usually considered to possess an additional symmetry, *conformal symmetry*, making it a 2-dimensional conformal field theory.

[30]Friedman (e.g., 2001, Lecture II, and Chapter 4).

[31]Examples include the "prediction" of a minimal length in LQG (Rovelli, 2004), and dimensional reduction in causal dynamical triangulations and HořavaLifshitz gravity, see Carlip (2017); 't Hooft (1993).

[32]E.g., Rovelli (2004).

we need to know what would count as a potential theory of QG. We need to know what QG is supposed to do, and what it is supposed to look like—*i.e., what constraints it is supposed to satisfy*. We need a more precise, generally accepted definition of QG, beyond the minimal characterisation offered above. Settling the constraints (role 3.) is, to my mind, an essential missing element of communicative rationality (although it is unclear whether such a definition would come as a result of communicative rationality, or whether it is a precursor to the activity—also, it is unclear whether these constraints must specifically be constitutive principles, but I suggest they need not be). As it stands, there are almost as many different conceptions of QG as there are approaches.[33]

4   Examples

As explained above, there are many different principles that play a role in the search for QG. Here, I present only a small selection, and do so very briefly. Nevertheless, my choice of principles here is suggestive of the sheer diversity of things being used as principles in QG—some of these are more precise than others, and some have origins closer to the empirical realm than others. Key principles such as Lorentz invariance, and unitarity, for example, are on the less-vague, closer-to-directly-observable sides of these two scales. Yet, as I illustrate with the examples, many of the central principles are vague, and they are often elevated not from experimental observations (or, if so, only via an indirect path), but from theoretical, or even heuristic results.

4.1   The (generalised) correspondence principle

What is it?   The requirement that any new theory "recover" or "reduce to" the (successful parts of the) incumbent theory in the domains where the latter is known to be successful.[34] Typically, this is supposed to be achieved via an appropriate limiting relation involving the characteristic constants of the new theory—but, while this may be sufficient, it is not necessary. In fact, correspondence relations between two theories are usually many and various, and may go in both directions.[35] So, in its broadest form, "correspondence" refers to any relationships that link two theories in the domains where both apply (i.e., provide successful descriptions), that establish (or are taken to establish, or are intended to establish) the *compatibility* of the two theories within these domains—i.e., that both theories approximately (i.e., including a small degree of error) have the same results in these domains. "Results" may include theoretical propositions as well as observational ones, and even "larger structures" such as derivations and explanations.

---

[33]For general reviews outlining the main approaches to QG, see Carlip (2001); Kiefer (2006); Rickles (2008a). It is also worthwhile to consult Butterfield & Isham (2001).

[34]The idea is essentially what Nickles (1973) calls "reduction$_2$".

[35]See, e.g., Hartmann (2002); Radder (1991); Post (1971).

Where does it come from?   The principle is usually attributed to Bohr, whose use of "the correspondence principle" was central to the development of old quantum theory, although what we now think of as the correspondence principle is not what Bohr meant by it.[36] Nevertheless, historically, the principle has featured heavily in the development of and acceptance of all new physical theories (i.e., it serves in all roles, 1.–5.). Its ubiquity is no doubt due to the perception of it as an "obvious requirement"—and if it is seen simply as the statement that the new theory should not conflict with the successful results of the old one, then it is. However, *the way* in which this agreement should be demonstrated is very much not obvious—for instance, how much of the older theory needs to be recovered, and what relationships are sufficient.

How is it used in QG?   The priciple is used heavily in all roles in theory construction and justification in QG; most conspicuously, it is a definitional constraint (3.i.), whose claim regards the application of 3.ii (i.e., it specifies particular external constraints), and whose satisfaction forms one of the few generally-agreed upon means of confirmation (4.i.) of the theory. Usually, the principle is explicitly presented as the justificatory requirement that QG recover GR in the classical limit—Carlip (2001, p. 927), for instance, refers to this as the "zeroth test" of QG. But, a fully satisfactory demonstration of the recovery of GR from QG requires us to bridge both the quantum/classical transition, by which we recover the classical appearance of spacetime (e.g., by decoherence, and a solution to the measurement problem), as well as the micro/macro transition, by which we return to familiar energy scales (e.g., via coarse graining, or a continuum limit, etc.).[37]

For approaches such as causal set theory, the *central goal* is to find a discrete, background independent theory that naturally reproduces GR spacetime in the appropriate domain, and thus the principle serves in roles (1.–4.).[38] Notable examples where correspondences with current theories have been presented as means of confirmation (4.) include the demonstration of the background spacetime of string theory satisfying the Einstein field equations in an appropriate limit, and the recovery, in causal set theory, of the correct spacetime dimension in accordance with GR. And, it is not just correspondence with the results of current theories *as they stand*, that serves to increase credence in QG, but also the recovery of theoretical results of combinations of current theories—e.g., the automatic appearance of the graviton (i.e., a massless spin-2 field) in the string spectrum is promoted in role 4.ii. by string theory (in spite of it being merely a hypothetical particle, unrecognised by either current framework), and the derivation of the Bekenstein-Hawking entropy in string theory and LQG features in role 4.i. (discussed below, §4.4).

Correspondence also serves important heuristic roles, as a guiding principle (1.);

---

[36]See, Bokulich (2014).

[37]See, Butterfield & Isham (1999, 2001); Wüthrich (2017).

[38]Causal set theory sees the desideratum of "being quantum" in the same way that string theory views the principle of "being background indpendent", i.e., an ultimate aspiration, but one that need not be implemented at the initial stages.

this form of correspondence works "bottom-up" (in energy scale), starting with the established theories, and attempting to infer parts of the new theory. In QG, an example is the number of approaches that take GR as a starting point, including all quantised-GR approaches (such as LQG), causal set theory, and causal dynamical triangulations. As stated, approaches that proceed this way should have the benefit of a solid foundation (agreed upon by all parties), and are able to appeal to this as means of justifying their theories—this is part of *prospective* communicative rationality. Thus, the generalised correspondence principle appears on the third level of Friedman's view of science. Yet, as I have attempted to demonstrate, its role is not, at this stage, part of *retrospective* communicative rationality—such a function is *post hoc*, able to be initiated only once the new theory is known, and would preclude the use of the principle in roles 1.–4., in the construction and justification of the theory-in-development.

4.2 UV-completion

What is it? A theory is said to be UV-complete if it is formally predictive to all possible high energy scales (or, equivalently, all short distance scales). In other words, the theory can be used to generate results that are potentially physically meaningful at all possible short distance scales—it does not guarantee that all of these results are (even approximately) *correct*, however (Newtonian mechanics is UV complete, for instance, since it does not formally 'break down' at any short distance scale, and yet we know that it is not the correct theory to use at all short distance scales).

The most familiar notion of a UV-complete theory is a QFT that is non-perturbatively (i.e., fixed-point) *renormalisable*, such as quantum chromodynamics.[39] Yet, the idea of UV-completion is not restricted to QFT, and there are other ways in which a theory may be UV-complete, apart from being renormalisable: i.e., UV-completion by cutoff, classicalisation, or UV/IR-correspondence.[40]

Where does it come from? (a.) In the heyday of QFT, UV-completion was usefully employed as a criterion of theory success: i.e., a means of selecting theories that were not only viable, but *physical*. While this changed with the development of the *renormalisation group* and the framework of *effective field theory*, together with the acceptance of its associated philosophy—that UV-*in*complete theories can also be physical, though they apply only within a restricted domain—the general perception of UV-incomplete theories as being faulty, or mathematically inconsistent, still lingers.[41] (b.) UV-completion appears to be a necessary condition for a theory to be *fundamental*, though note that QG need not be a fundamental theory (Crowther & Linnemann, 2017). (c.) A theory of QG may be UV-complete if it describes a minimal length

---

[39]*Perturbative* renormalisability is insufficient to establish UV-completion, since a theory may be perturbatively renormalisable may still face a Landau Pole, as is the case in quantum electrodynamics.

[40]Butterfield & Bouatta (2015); Crowther & Linnemann (2017); Dvali et al. (2011); Zee (2010).

[41]Cf. Cao & Schweber (1993); Shankar (1999); Weinberg (1979, 1999).

scale, supposedly representing the shortest possible scale that can be probed in principle. There are various arguments for such a minimal length (Hossenfelder, 2013). (d.) Tied up with these motivations is the fact that one natural way of combining GR and QFT (treating GR in the framework of perturbative QFT, henceforth "perturbative quantum GR") produces a theory that is non-renormalisable—and thus UV-incomplete—breaking down at the Planck scale. This leads to the theory being widely (though mistakenly) viewed as inconsistent, and drives the search for a UV-complete theory of QG (when all that is actually required in this respect is a "UV-better" theory, predictive at the Planck scale).[42]

How is it used in QG? UV-completion is used as a central guiding principle in the search for QG, and has been (mis)interpreted as principle defining what would count as a successful theory, as researchers seek to overcome the problem of the non-renormalisability of perturbative quantum GR—although, as stated above, a UV-complete theory is not in fact necessary to solve this problem. The desire to obtain a UV-complete theory is actually the *primary* motivation driving a number of the approaches to QG, including asymptotic safety (Niedermaier & Reuter, 2006), and higher-derivative approaches (Kiefer, 2007; Stelle, 1977). Additionally, in several of the approaches where UV-completion appears to have been obtained (or where it appears it could be obtained), this is presented as evidence in favour of these approaches by their proponents. This is the case, for example, in LQG, where UV-completion is not taken as a guiding principle, but is nevertheless automatically obtained through quantisation and a minimal length (Rovelli, 2004). UV-completion is also obtained and presented as a means of confirmation (role 4) in alternative theories of gravity (where, again, it is not used necessarily as a central motivation), for instance Brans-Dicke and Hořava-Lifshitz gravity.[43]

String theory is supposed to be a "theory of everything", rather than just QG—as such, it must be UV-complete. The extendedness of the theory's basic entities is very likely to ensure that it is UV-complete, although no proof has yet been found that it is (see Hagar (2014, §7.2), Dawid (2013, chapter 1)). Textbooks on string theory often present an analogy between the perturbative non-renormalisability of quantum GR and that of 4-Fermi theory, which is a non-renormalisable theory that was revealed to be the low-energy limit of the *renormalisable* theory of electroweak interactions. Similiarly, proponents of string theory claim that string theory is the renormalisable theory underlying perturbative quantum GR. Thus, the alleged UV-completeness of string theory is presented as one of its selling points (i.e., as a criterion of justification).

Note that, in some approaches to QG, UV-completion may conflict with other principles that are desirable and otherwise apparently viable. For instance, if UV-completion is obtained by cutoff, it may pose problems for Lorentz invariance, and, in

---

[42]Cf. Crowther & Linnemann (2017).

[43]For Brans-Dicke gravity, see, Haba (2002) *contra* Deser & van Nieuwenhuizen (1974) *and* 't Hooft & Veltman (1974); for Hořava-Lifshitz gravity, see Orlando & Reffert (2009).

the higher-derivative approaches UV-completion apparently conflicts with the principle of unitarity. Crowther & Linnemann (2017) also argues that UV-completion may conflict with the goal of unification.

4.3 Background independence

What is it? The general idea is (a.) that QG should not feature a fixed (background) spacetime, but there are several meanings, corresponding to different notions of "fixity" (Butterfield & Isham, 1999, §4.1), as well as the different (layers of) structures that could be identified as "spacetime" (e.g., metric, manifold, affine structure, etc.). There are (at least) four proposals that attempt to formally capture the idea of background independence. I list these definitions concisely (and without the necessary elaboration), but note that each is a separate proposal: (i) A theory is background independent if it features no absolute objects (i.e., structures that are the same in all models of the theory); (ii) A theory is background independent if it has no formulation which features fixed fields (i.e., there is no formulation of the theory that is not diffeomorphism invariant); (iii) A theory is background independent if its solution space is determined by a generally-covariant action (all of whose dependent variables are subject to Hamilton's principle, and represent physical fields) (iv) A theory is background independent if all of its physical degrees of freedom correspond to geometrical degrees of freedom.[44] Each of these attempted characterisations has its own difficulties, however—and, arguably, none are adequate to do the job.[45]

Two other common interpretations of the principle are (b.) that the theory be *generally covariant*, and (c.) that it be *non-perturbative* (a perturbative treatment of gravity involves approximating a dynamical spacetime by splitting the metric into a fixed background part, plus a small disturbance on it).

Where does it come from? (a.) The broad desire for background independence is driven by a number of familiar sentiments in the philosophy of space and time, including, e.g., the dissatisfaction with structures that need to be specified in our theories, rather than determined by them (e.g., such structures apparently stand in need of explanation), the suspicion of absolute structures (e.g., recalling absolute space in Newtonian mechanics as something that acts but cannot be acted upon), and the idea that physics progresses by seeking ever more relational theories (Smolin, 2006).

(b.) General covariance was taken as a strong constraint and guiding principle (roles 3. and 1.) by Einstein in developing GR. This form of the principle may be thought of as the result of pushing to the extreme the idea that physics be independent of the coordinate system used to describe it (i.e., no preferred reference frame). Yet, in 1917, Kretschmann showed that the very idea of general covariance may be trivial, by claiming that any theory can be made covariant, through the addition of non-

---

[44]For details on (i–iii) see Pooley (2017), and for (iv), see Belot (2011).

[45]Further details: Giulini (2007); Pitts (2006); Pooley (2017); Read (2016).

dynamical auxiliary fields. Thus, Friedman (1983, p. 55) claims that "the principle of general covariance has no physical content whatever: it specifies no particular physical theory; rather it merely expresses our commitment to a certain style of formulating physical theories." Einstein, however, maintained that general covariance could still be used as a principle of theory-selection, if it is paired with a conception of "formal simplicity"—while any theory may be made generally-covariant, only those expressions which are formally simple after having been written in a generally-covariant form should be considered as candidates for physical laws. Still, as Giulini (2007) points out, a generally covariant form of Newtonian gravity (given by Cartan and Friederichs) is still not outrageously complicated (though perhaps somewhat unnatural).

While it is not impossible that general covariance be used in defining "what is special" about GR,[46] and thus feature in a principle of theory selection, general covariance can not *by itself* be the principle of background independence. Most of the recent literature has, though, instead moved on in to proposals (a.) in their attempts to salvage the general intuition that GR is special by virtue of possessing some particular conception of background independence. In this context, the proposal (i) above, fails for GR (and any theory with massive fields),[47] and while proposal (ii) arguably works for GR, it is difficult to implement as a guiding principle in the search for a new theory (Read, 2016). Nevertheless, the success of GR as a theory of spacetime, together with the significant heuristic role of the principle in its discovery, mean, however, that background independence (in some sense, which may, arguably, be any of (ii-iv)) continues to be viewed as a desirable principle in QG, in all roles.[48] And note that even the principle of general covariance, and (i) can plausibly serve heuristic roles.

Finally, the distinct notion, (c): The desire that a theory be non-perturbative is a natural one, given that perturbation theory is a set of techniques used to construct approximations when the exact theory is unknown or unsolvable. These techniques are of limited applicability, and can easily lead to problems when misapplied.

How is it used in QG?   Currently, string theory is considered a background *dependent* approach: it is based on perturbative methods, and describes strings moving on a background spacetime—and thus, it apparently fails on all (a.–c.). This is widely recognised as a deficiency of the approach. But, as Read (2016, §6) demonstrates (building on a claim made by Huggett & Vistarini (2015)), there are in fact several senses in which string theory (considered both at the level of spacetime fields, and at the level of worldsheet fields) may be considered background *independent*, on some of the definitions (ii–iv) above, given that the background fields in the theory are required

---

[46]Cf. Norton (2003); also Brown (2005, §5.3.1).

[47]See references in Fn. 45.

[48]However, given the above discussion, one might wonder whether the principle should be viewed as being so significant—a thought that gains further weight with the recognition that particle physicists in the 1930s–70s managed to derive Einstein's equations of GR from entirely unrelated principles, including universal coupling and avoidance of ghosts (e.g., Desser, 1970; Van Nieuwenhuizen, 1973). (Thanks to a referee for offering this food for thought).

to be dynamically coupled together in the same dynamical equations of GR (the Einstein field equations, plus higher order corrections), and thus are not "fixed" after all, but possess background independence in ways similar to GR.

Additionally, there are attempts to construct a non-perturbative version of string theory; the most celebrated development being the *AdS/CFT duality* (also known as the *Maldacena conjecture*, after Maldacena (1998)). This is a relationship—an exact equivalence[49]—between a string theory featuring gravity, describing closed strings propagating on a spacetime (anti-de Sitter space (AdS)), known as the "bulk", and a gauge theory without gravity (a conformal field theory (CFT)), defined on the boundary that contains the bulk spacetime. In the regime where the string theory is *strongly coupled*, necessitating the use of perturbative methods—the CFT is *weakly coupled*, meaning that it can be defined non-perturbatively. Since the theories are supposed to be equivalent, the non-perturbative regime of the CFT can be used to shed light on the perturbative regime of the string theory.[50] (This is closely related to the holographic principle, §4.4).

While string theorists hope to ultimately arrive at a theory that is background independent, they started out with a background dependent approach. So, the principle features as a fallible constraint (5.)—a general desideratum, such that, credence in the approach increases by its being shown to be background independent (role 4.) Eventually, it may also feature as a criterion of acceptance (2.).

Other approaches to QG, including, e.g., causal set theory, LQG, and causal dynamical triangulations, use the principle in all roles (1.–4., and role 5. is not required). These approaches strive for a fully background independent theory, and sell themselves as upholding what they see as a central insight of Einstein gravity (and of any acceptable theory). LQG, in particular, is promoted as continuing the tradition of increasing *relationalism* in spacetime theories (see, e.g., Smolin, 2006; Rovelli, 2004). Rather than give examples of all these roles of background independence, I just mention the interesting case of its featuring as an unexpected means of confirmation (4.ii.) in the early development of LQG. As Rovelli (2004, pp. 11–12) describes, attempts to formulate a theory with loop states (the fundamental kinematic states described by LQG) on a continuous background (as a QFT) were unsuccessful: because an infinitesimal displacement of a loop state against the background produced a new, distinct, loop state, the theory yielded a continuum of loop states. The resulting space spanned by these states was thus "too big" to provide the basis of a separable Hilbert space. This problem was solved, however, by moving to a background indpendent theory—defining the position of a loop *relationally* (i.e., with respect to other loops), dramatically reduces the number of distinct loop states (an implementation of diffeomorphism invariance), and the theory becomes viable. The unexpected "success" of background independence is thus taken as a signal that the approach is on the right track (4.ii.).

---

[49]The meaning of this is explored by, e.g., Butterfield (Forthcoming); Dawid (2017); de Haro (2017); Read & Møller-Nielsen (2018); Rickles (2011); Teh (2013).

[50]See, e.g., Polchinski (2017).

4.4 The holographic principle

What is it? The most common understanding of the principle is based on the idea that the description of a volume—"bulk"—of space can be equivalently defined on a surface bounding the volume of space (the boundary has a lower dimension than the bulk it contains). In other words, a physical theory defined only on the (N-1)-dimensional boundary of the region it encloses completely describes the N-dimensional physics of the bulk. The principle is then the claim that the full theory of QG can be reformulated as a theory all of whose degrees of freedom are defined on the boundary—a concrete example is the AdS/CFT duality (outlined above, §4.3).

This is closely related to the idea of an *entropy bound*—in essence, the statement that there is a limit to the amount of entropy (typically, though not necessarily, understood as *information*) that can exist within a given region of spacetime, and that this limit depends on the area of the surface bounding the region, rather than the volume of the region. Various entropy bounds have been proposed, for instance, the *spherical entropy bound*, which states that the entropy S of a matter system is, $S \leq A/4$, where A is a suitably defined area enclosing the matter system (Susskind, 1995). The first proposed was the *Bekenstein bound* (Bekenstein, 1981), and the current most general form is the *covariant entropy bound*, (CEB), which is framed in terms of constructions called *light sheets* (Bousso, 2002). But Smolin (2001) argues that only a different form, the *weak holographic principle*, is capable of surviving in a full quantum theory of gravity (specifically, Smolin argues that this is the only form that is compatible with the other principles that should feature in QG, including background independence).

Where does it come from? The principle originates in *black hole thermodynamics* (BHT), which is the proposal that black holes are thermodynamical objects, possessing entropy proportional to their (event horizon) area, and satisfying a set of laws analogous to those of ordinary (i.e., matter) thermodynamics. This proposal itself was motivated by the worry that black holes may otherwise violate the second law of thermodynamics, and gained widespread acceptance after Hawking's (1975) calculation showing that black holes radiate as thermodynamic objects. Hawking's calculation set the proportionality constant in Bekenstein's entropy equation, so that the Bekenstein-Hawking entropy of a black hole is accepted as, $S_{BH} = kc^3 A_{BH}/4\hbar G$, where A_{BH} is the are of the black hole's horizon, and k is Boltzmann's constant. Yet, the whole enterprise of black hole thermodynamics is entirely theoretical—based on a strong network of *semiclassical* results rather than any empirical observations or experimental tests.[51] Historically, the idea of the holographic principle was tied, in part, to the debate over the information loss paradox (See, e.g., Bigatti & Susskind, 2000). But, authors such as Bousso (2002) emphasise that, although this may be the historical route that led to the discovery of the entropy bound, the holographic principle itself does not depend on the correctness of BHT, and nor does BHT *explain* the existence of the entropy bound.

---

[51]Many philosophers have thus recommended caution in interpreting these results as physically meaningful, but cf. Wallace (2017a,b), which take a firmer stance.

How is it used in QG? We should distinguish, firstly, between BHT—including the Bekenstein-Hawking entropy result—and, secondly, the holographic principle that it inspired. First: entropy is linked to the number of *micro states* of a system, and its connection with black holes (purely gravitational objects) prompts the search for analogous micro states underlying GR; additionally, it provides a measure of the number of such micro states, which QG is expected to recover and explain. Thus, the Bekenstein-Hawking result is generally treated as a guiding principle (1.), a criterion of acceptance (3.), and a means of confirmation (4.). Both string theory and LQG have (in their own ways) managed to reproduce the value of the Bekenstein-Hawking entropy under certain conditions and for particular types of black holes (neither have achieved a general result).

Second: the holographic principle (along with the existence of an entropy bound) is treated as a striking result inexplicable according to current theories (where we generally expect entropy to scale with volume rather than area), and is thus taken as indication of the deficiency in current physics, i.e., it is a principle that motivates the search for QG.[52] The entropy bound itself, however, has been even more elusive. Rideout & Zohren (2006) show how the spherical entropy bound can be obtained using causal set theory (demonstrating its use in roles 1., 3. and 4. in that approach). In string theory, the holographic principle has been proposed as a guiding principle: Bousso (2002, §IX.A.) suggests that, since the covariant entropy bound is valid in highly dynamical spacetimes, and string theory has had difficulty in describing such geometries, that (this form of) the holographic principle could aid further development of the theory. Bousso also presents the holographic principle as an unexpected means of confirmation (4.ii.) of string theory. The holographic principle is only expected to be fully manifest in a non-perturbative formulation of the theory, and, in the years following the proposal of the holographic principle, non-perturbative descriptions of string theory have been found for two special cases—one of which is the AdS/CFT duality. As mentioned above, this is a concrete manifestation of the principle, which thus serves to increase credence in string theory.

The principle may also be taken as a postulate (2.); Smolin (2017) suggests using a form of the holographic principle in the attempt to build QG as a principle theory, in Einstein's sense. Also, the holographic principle could potentially function as a constitutive principle—indeed, it seems the most likely candidate for a coordinating principle defining a new physical framework. For instance, Bousso (2002, p. 861) suggests that the holographic principle may not only aid the search for a non-perturbative version of string theory, but "could also contribute to a background-independent formulation that would illuminate the conceptual foundation of string theory". Smolin (2001, p. 210) also presents his weak holographic as playing something like a coordinating role,

---

[52]Some work has been done in investigating the area scaling law. Oppenheim (2003), for example, uses an analogue model of a black hole with long-range interactions and shows how the entropy goes from scaling with volume to scaling with area as the strength of the interactions increases. Chandran et al. (2016) recounts the applicability of the area law for *entanglement entropy*, showing under what conditions it scales differently.

for instance, he states that,

> [T]he weak holographic principle enters a background independent quantum theory of cosmology as a framework for that theory's interpretation and measurement theory. Its role is to constrain the quantum causal structure of a quantum spacetime in a way that connects the geometry of the surfaces on which measurements may be made with a measure of the information that those measurements may produce. In this context the entropy bound becomes a definition, by which the notion of geometry is reduced to more fundamental notions coming from the quantum theory of cosmology.

5   Conclusion

In this paper, I aimed to illustrate the sheer diversity of principles featuring in the search for QG—many of which may not be recognised as principles on particular philosophical understandings of the term, but which nevertheless play key roles in motivating the need for the new theory, as well as defining it. In particular, I distinguished five different roles that principles play in theory construction and justification in QG, accompanied by numerous examples, including four whose form, origins and roles, were outlined in some detail. These are the generalised correspondence principle, UV-completion, background independence, and the holographic principle—the last of which is currently the most promising candidate for a new "constitutive principle" (i.e., a coordinative definition) that could provide the basis for a new framework of physics, yet it comes not from the empirical realm, but from striking theoretical results.

I presented the principles neutrally, but also suggested two other useful perspectives from which to view the crisis. Firstly, through the immediate perspective of the researchers involved—in a way analogous to that of Poincaré and Einstein during the crises at the turn of the 20th century—where the problem involves assessing the damage on current principles, and settling the constraints on the new theory, as well as choosing the appropriate guiding principles and postulates to lead you to the theory that satisfies these constraints. I also suggested viewing the crisis through Friedman's picture of scientific theories; here, I argued that the crisis is a clash of principles that apparently suggest the criticality of a revolution. This conflict is the result of current physics, as a whole, being based on an incoherent framework in Friedman's sense.

Yet, I argued that a revolution in Friedman's sense may not be necessary, and that several of the current approaches to QG, including the most prominent contender, are at present non-revolutionary—they are not based on new constitutive principles, and do not represent a new framework in Friedman's sense. Instead, they deal with the conflict through the neglect of particular principles, and build their approaches on parts of the current framework at the expense of others. This, I argued, means that communicative rationality is fractured. Thus, although the situation in QG research currently does not reflect Friedman's views, these may be used to suggest new ways for QG to move forward—for instance, by mending communicative rationality (through, I

argue, the development of the generalised correspondence principle as a more precise constraint on the theory); and by seeking new constitutive principles. Additionally, if we translate Friedman's thesis into a prescriptive strategy, it also points us to look for the new constitutive principles in a direction that has so far been neglected—to established, mundane empirical facts, rather than novel and striking theoretical results.

Acknowledgements

References

Almheiri, A., Marolf, D., Polchinski, J., & Sully, J. (2013). Black holes: complementarity or firewalls? *Journal of High Energy Physics*, *2013*(2), 62.

Anderson, E. (2012). Problem of time in quantum gravity. *Annalen der Physik*, *524*(12), 757–786.

Bekenstein, J. (1981). A universal upper bound on the entropy to energy ratio for bounded systems. *Phys. Rev. D*, *23*(2), 287.

Belot, G. (2011). Background-independence. *General Relativity and Gravitation*, *43*(10), 2865–2884.

Belot, G. & Earman, J. (2001). Presocratic quantum gravity. In C. Callender & N. Huggett (Eds.), *Physics Meets Philosophy at the Planck Scale* (pp. 213–255). Cambridge: Cambridge University Press.

Belot, G., Earman, J., & Ruetsche, L. (1999). The hawking information loss paradox: The anatomy of a controversy. *The British Journal for the Philosophy of Science*, *50*(2), 189–229.

Bergmann, P. G. & Komar, A. B. (1960). Poisson brackets between locally defined observables in general relativity. *Phys. Rev. Lett.*, *4*, 432–433.

Bigatti, D. & Susskind, L. (2000). The holographic principle. In L. Thorlacius & T. Jonsson (Eds.), *M-Theory and Quantum Geometry*, volume 556 of *NATO Science Series* (pp. 179–226). Dordrecht: Kluwer Academic Publishers.

Blum, A. & Rickles, D. (Eds.). (Forthcoming). *Quantum Gravity in the First Half of the Twentieth Century: A Sourcebook*. Berlin: Edition Open Access.

Bokulich, A. (2014). Bohr's correspondence principle. *The Stanford Encyclopedia of Philosophy*, *Spring 2014*.

Bousso, R. (2002). The holographic principle. *Reviews of Modern Physics, 74*, 825–874.

Brown, H. (2005). *Physical relativity: space-time structure from a dynamical perspective*. Oxford: Oxford University Press.

Butterfield, J. (Forthcoming). On dualities and equivalences between physical theories. In N. Huggett, B. L. Bihan, & C. Wüthrich (Eds.), *Philosophy Beyond Spacetime*. Oxford: Oxford University Press. *https://arxiv.org/abs/1806.01505*.

Butterfield, J. & Bouatta, N. (2015). Renormalization for philosophers. In *Metaphysics in Contemporary Physics* (pp. 437–485). Leiden: Brill.

Butterfield, J. & Isham, C. (1999). On the emergence of time in quantum gravity. In J. Butterfield (Ed.), *The Arguments of Time* (pp. 116–168). Oxford: Oxford University Press.

Butterfield, J. & Isham, C. (2001). Spacetime and the philosophical challenge of quantum gravity. In C. Callender & N. Huggett (Eds.), *Physics Meets Philosophy at the Planck Scale* (pp. 33–89). Cambridge: Cambridge University Press.

Callender, C. & Huggett, N. (Eds.). (2001). *Physics Meets Philosophy at the Planck Scale: Contemporary Theories in Quantum Gravity*. Cambridge: Cambridge University Press.

Cao, T. Y. & Schweber, S. S. (1993). The conceptual foundations and the philosophical aspects of renormalization theory. *Synthese, 97*(1), 33–108.

Carlip, S. (2001). Quantum gravity: a progress report. *Reports on Progress in Physics, 64*(8), 885.

Carlip, S. (2017). Dimension and dimensional reduction in quantum gravity. *Classical and Quantum Gravity, 34*(19), 193001.

Chandran, A., Laumann, C., & Sorkin, R. (2016). When is an area law not an area law? *Entropy, 18*(7), 240–247.

Crowther, K. (2016). *Effective Spacetime: Understanding Emergence in Effective Field Theory and Quantum Gravity*. Heidelberg: Springer.

Crowther, K. & Linnemann, N. (2017). Renormalizability, fundamentality and a final theory: The role of UV completion in the search for quantum gravity. *British Journal for the Philosophy of Science*.

Crowther, K. & Rickles, D. (2014). Introduction: Principles of quantum gravity. *Studies In History and Philosophy of Modern Physics, 46*, 135–141.

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

Dawid, R. (2017). String dualities and empirical equivalence. *Studies In History and Philosophy of Modern Physics*, *59*, 21–29.

de Haro, S. (2017). Dualities and emergent gravity: Gauge/gravity duality. *Studies In History and Philosophy of Modern Physics*, *59*, 109–125.

Deser, S. & van Nieuwenhuizen, P. (1974). One-loop divergences of quantized Einstein-Maxwell fields. *Physical Review D*, *10*(2), 401.

Desser, S. (1970). Self-interaction and gauge invariance. *General Relativity and Gravitation*, *1*, 9–18.

Dvali, G., Giudice, G. F., Gomez, C., & Kehagias, A. (2011). Uv-completion by classicalization. *Journal of High Energy Physics*, *2011*(8), 1–31.

Earman, J. (2006). The implications of general covariance for the ontology and ideology of spacetime. In D. Dieks (Ed.), *The Ontology of Spacetime* (pp. 3–23). Amsterdam: Elsevier.

Einstein, A. (1916). Approximate integration of the field equations of gravitation. In *The Collected Papers of Albert Einstein*, volume Volume 6: The Berlin Years: Writings, 1914-1917 (English translation supplement). http://einsteinpapers.press.princeton.edu/vol6-trans/213.

Einstein, A. (1919). Time, space, and gravitation. *The Times, London, November 28*, 13–14.

Friedman, M. (1983). *Foundations of Space-Time Theories*. Princeton: Princeton University Press.

Friedman, M. (2001). *Dynamics of Reason: The 1999 Kant Lectures of Stanford University*. Stanford: CSLI Publications.

Giulini, D. (2007). Remarks on the notions of general covariance and background independence. In I.-O. Stamatescu & E. Seiler (Eds.), *Lecture Notes in Physics*, volume 721 (pp. 105–120). Berlin: Springer.

Haba, Z. (2002). Renormalization in quantum brans-dicke gravity. *arXiv preprint hep-th/0205130*.

Hagar, A. (2014). *Discrete or continuous? The quest for fundamental length in modern physics*. Cambridge: Cambridge University Press.

Hartmann, S. (2002). On correspondence. *Studies In History and Philosophy of Modern Physics*, *33*(1), 79–94.

Hawking, S. (1974). Black hole explosions? *Nature*, *248*, 30–31.

Hawking, S. (1975). Particle creation by black holes. *Communications in Mathematical Physics*, *43*, 199–220.

Hossenfelder, S. (2013). Minimal length scale scenarios for quantum gravity. *Living Reviews in Relativity*, (2).

Huggett, N. & Callender, C. (2001). Why quantize gravity (or any other field for that matter)? *Philosophy of Science*, *68*(3), S382–S394.

Huggett, N. & Vistarini, T. (2015). Deriving general relativity from string theory. *Philosophy of Science*, *82*(5), 1163–1174.

Huggett, N., Vistarini, T., & Wüthrich, C. (2013). Time in quantum gravity. In H. Dyke & A. Bardon (Eds.), *A Companion to the Philosophy of Time*, Blackwell Companions to Philosophy (pp. 242–261). Chichester: Wiley-Blackwell.

Huggett, N. & Wüthrich, C. (2013). Emergent spacetime and empirical (in)coherence. *Studies in History and Philosophy of Modern Physics*, *44*(3), 276–285.

Kiefer, C. (2006). Quantum gravity: General introduction and recent developments. *Annals of Physics*, *15*(1), 129–148.

Kiefer, C. (2007). Why quantum gravity? In *Approaches to Fundamental Physics* (pp. 123–130). Springer.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Maldacena, J. (1998). The large N limit of superconformal field theories and supergravity. *Advances in Theoretical and Mathematical Physics*, *2*, 231–252.

Mathur, S. (2009). The information paradox: a pedagogical introduction. *Classical and Quantum Gravity*, *26*, 224001.

Mattingly, J. (2005). Is quantum gravity necessary? In A. Kox & J. Eisenstaedt (Eds.), *The Universe of General Relativity* (pp. 327–338). Birkhäuser.

Mattingly, J. (2006). Why Eppley and Hannah's thought experiment fails. *Phys. Rev. D*, *73*, 062025.

Mattingly, J. (2009). Mongrel gravity. *Erkenntnis*, *70*(3), 379–395.

Nickles, T. (1973). Two concepts of intertheoretic reduction. *The Journal of Philosophy*, *70*(7), 181–201.

Niedermaier, M. & Reuter, M. (2006). The asymptotic safety scenario in quantum gravity. *Living Rev. Rel*, *9*(5), 173.

Norton, J. (2003). General covariance, gauge theories and the Kretschmann objection. In K. Brading & E. Castellani (Eds.), *Symmetries in Physics: Philosophical Reflections* (pp. 110–123). Cambridge University Press.

Oppenheim, J. (2003). Thermodynamics with long-range interactions: From Ising models to black holes. *Physical Review E*, (68), 016108.

Orlando, D. & Reffert, S. (2009). The renormalizability of Hořava-Lifshitz-type gravities. *Classical and Quantum Gravity*, *26*(15), 155021.

Percacci, R. (2009). Asymptotic safety. In D. Oriti (Ed.), *Approaches to Quantum Gravity: Towards a New Understanding of Space, Time and Matter* (pp. 111–128). Cambridge: Cambridge University Press.

Pitts, J. B. (2006). Absolute objects and counterexamples: Jones-Geroch dust, Torretti constant curvature, tetrad-spinor, and scalar density. *Studies In History and Philosophy of Modern Physics*, *37*, 347–371.

Pitts, J. B. (2014). Change in Hamiltonian general relativity from the lack of a time-like killing vector field. *Studies in History and Philosophy of Modern Physics*, *47*, 68 – 89.

Pitts, J. B. (2017). Equivalent theories redefine Hamiltonian observables to exhibit change in general relativity. *Classical and Quantum Gravity*, *34*(5), 055008.

Poincaré, H. (1905a). The principles of mathematical physics. *The Monist*, *15*(1), 1–24.

Poincaré, H. (1905b). *Science and Hypothesis*. New York: Walter Scott.

Poincaré, H. (1907). *The Value of Science*. New York: Science Press.

Polchinski, J. (2017). Dualities of fields and strings. *Studies In History and Philosophy of Modern Physics*, *59*, 6–20.

Pons, J., Salisbury, D., & Sundermeyer, K. (2010). Observables in classical canonical gravity: Folklore demystified. *Journal of Physics: Conference Series*, *222*(1), 012018.

Pons, J. M. (2005). On Dirac's incomplete analysis of gauge transformations. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *36*(3), 491 – 518.

Pooley, O. (2015). Background independence, diffeomorphism invariance, and the meaning of coordinates. In D. Lehmkuhl (Ed.), *Towards a Theory of Spacetime Theories* (pp. 105–143). Birkhäuser.

Pooley, O. (2017). Background independence, diffeomorphism invariance, and the meaning of coordinates. In *Towards a Theory of Spacetime Theories* (pp. 105–144). Birkhäuser.

Post, H. (1971). Correspondence, invariance and heuristics: In praise of conservative induction. *Studies in History and Philosophy of Science Part A, 2*(3), 213–255.

Radder, H. (1991). Heuristics and the generalized correspondence principle. *British Journal for the Philosophy of Science, 42*, 195–226.

Read, J. (2016). Background independence in classical and quantum gravity. Master's thesis, University of Oxford.

Read, J. & Møller-Nielsen, T. (2018). Motivating dualities. *Synthese, https://doi.org/10.1007/s11229-018-1817-5.*

Rickles, D. (2006a). Time and structure in canonical gravity. In D. Rickles, S. French, & J. Saatsi (Eds.), *The Structural Foundations of Quantum Gravity* (pp. 152–196). Oxford: Oxford University Press.

Rickles, D. (2006b). Who's afraid of background independence? In D. Dieks (Ed.), *The Ontology of Spacetime* (pp. 133–152). Amsterdam: Elsevier.

Rickles, D. (2008a). Quantum gravity: A primer for philosophers. In D. Rickles (Ed.), *The Ashgate Companion to Contemporary Philosophy of Physics* chapter 5, (pp. 262–365). Aldershot: Ashgate.

Rickles, D. (2008b). *Symmetry, Structure, and Spacetime.* Amsterdam: Elsevier.

Rickles, D. (2011). A philosopher looks at string dualities. *Studies in History and Philosophy of Modern Physics, 42*(1), 54–67.

Rickles, D. (2012). Time, observables, and structure. In E. Landry & D. Rickles (Eds.), *Structural Realism*, volume 77 of *The Western Ontario Series in Philosophy of Science* (pp. 135–145). Dordrecht: Springer.

Rideout, D. & Zohren, S. (2006). Evidence for an entropy bound from fundamentally discrete gravity. *Classical and Quantum Gravity, 23*(22), 6195.

Rovelli, C. (1991). What is observable in classical and quantum gravity? *Classical and Quantum Gravity, 8*(2), 297.

Rovelli, C. (2002). GPS observables in general relativity. *Phys. Rev. D, 65*, 044017.

Rovelli, C. (2004). *Quantum Gravity.* Cambridge: Cambridge University Press.

Samaroo, R. (2015). Friedman's thesis. *Studies in History and Philosophy of Modern Physics, 52*, 129–138.

Shankar, R. (1999). The triumph and limitations of quantum field theory. In T. Y. Cao (Ed.), *Conceptual Foundations of Quantum Field Theory* (pp. 47–55). Cambridge: Cambridge University Press.

Smolin, L. (2001). The strong and weak holographic principles. *Nuclear Physics B*, *601*(12), 209 – 247.

Smolin, L. (2006). The case for background independence. In D. Rickles, S. French, & J. Saatsi (Eds.), *The Structural Foundations of Quantum Gravity* (pp. 196–239). Oxford: Oxford University Press.

Smolin, L. (2017). Four principles for quantum gravity. In J. Bagla & S. Engineer (Eds.), *Gravity and the Quantum*, volume 187 of *Fundamental Theories of Physics* (pp. 427–450). Springer.

Stelle, K. (1977). Renormalization of higher-derivative quantum gravity. *Physical Review D*, *16*(4), 953.

Susskind, L. (1995). The world as a hologram. *Journal of Mathematical Physics*, *36*(11), 6377–6396.

't Hooft, G. (1993). Dimensional reduction in quantum gravity. *https://arxiv.org/abs/gr-qc/9310026*.

't Hooft, G. & Veltman, M. (1974). One-loop divergencies in the theory of gravitation. In *Annales de l'IHP Physique théorique*, volume 20, (pp. 69–94).

Teh, N. (2013). Holography and emergence. *Studies In History and Philosophy of Modern Physics*, *44*(3), 300–311.

Thbault, K. P. (2012). Three denials of time in the interpretation of canonical gravity. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *43*(4), 277 – 294.

Van Nieuwenhuizen, P. (1973). On ghost-free tensor lagrangians and linearized gravitation. *Nucl. Phys. B*, (60), 478–492.

Wallace, D. (2017a). The case for black hole thermodynamics, part i: phenomenological thermodynamics. *https://arxiv.org/abs/1710.02724*.

Wallace, D. (2017b). The case for black hole thermodynamics, part ii: statistical mechanics. *https://arxiv.org/abs/1710.02725*.

Wallace, D. (2018). Why black hole information loss is paradoxical. *https://arxiv.org/abs/1710.03783v2*.

Weinberg, S. (1979). Ultraviolet divergencies in quantum theories of gravitation. In S. Hawking & W. Israel (Eds.), *General relativity, an Einstein Centenary survey* (pp. 790–831). Cambridge: Cambridge University Press.

Weinberg, S. (1999). What is quantum field theory and what did we think it was? In T. Y. Cao (Ed.), *Conceptual Foundations of Quantum Field Theory* (pp. 241–251). Cambridge University Press.

Wheeler, J. (1984). Quantum gravity: The question of measurement. In S. Christensen (Ed.), *Quantum Theory of Gravity: Essays in Honor of the 60th Birthday of Bryce S. DeWitt* (pp. 224–233). Bristol: Adam Hilger.

Wheeler, J. & Ford, K. (1998). *Geons, Black Holes and Quantum Foam*. W.W. Norton & Company.

Wüthrich, C. (2005). To quantize or not to quantize: Fact and folklore in quantum gravity. *Philosophy of Science, 72*, 777–788.

Wüthrich, C. (2017). Raiders of the lost spacetime. In D. Lehmkuhl (Ed.), *Towards a Theory of Spacetime Theories* (pp. 297–335). Birkhäuser.

Zee, A. (2010). *Quantum field theory in a nutshell*. Princeton University Press.