What's the problem with the cosmological constant?

Mike D. Schneider*†

Abstract

The "Cosmological Constant Problem" (CCP) is widely considered a crisis in contemporary theoretical physics. Unfortunately, the search for its resolution is hampered by open disagreement about what is, strictly, the problem. This disagreement stems from the observation that the CCP is not a problem within any of our current theories, and nearly all of the details of those future theories for which the CCP could be made a problem are up for grabs. Given this state of affairs, I discuss how one ought to make sense of the role of the CCP in physics and generalize some lessons from it.

1   Introduction

The "Cosmological Constant Problem" (CCP) has occupied the attention of theoretical physicists for several decades, often glamorized in popular science as "the worst prediction in the history of physics". According to INSPIRE, the High Energy Physics online information system, Weinberg's 1989 paper that first formally declared it as a crisis has been cited over 3900 times.

Despite all of this activity, what is actually meant by the CCP is ambiguous. In most cases, the CCP concerns a tension between the measured value assigned to the cosmological constant ("Λ") in the standard model of cosmology and the values of certain quantities predicted in quantum field theory (QFT) that are usually thought to influence it. But the status of the CCP as a problem is made complicated by two observations. First, it is unclear by what criteria the quantities from QFT are considered to be relevant to the standard model of cosmology built on classical (i.e. not quantum) physics. Second, even if such quantities were made relevant, the current standard model of cosmology can accommodate them via an unconstrained parameter, called the "bare cosmological constant" or, in context, the "bare term" (c.f. Carroll (2001); Martin (2012)), usually denoted as Λ_0. As such, the CCP cannot be understood as a problem within the current standard model of cosmology. (Section 2 will formally develop this claim.)

The situation is quite different when one leaves behind present, well-evidenced physical theory and speculates about what a quantum theory of gravity will entail for the future of cosmology. But competing semiclassical intuitions about what our current theories of general relativity (GR) and QFT imply about future theories of quantum gravity make it unclear how to assess that conflict: different assumptions about what

quantum gravity will ultimately entail give rise to different characterizations of the CCP. For instance, whether or not one believes that the future theory includes some feature responsible for a non-vanishing bare term Λ_0 in the classical gravitational limit influences what one expects ought to be so in regards to the matter limit of that same theory (in which gravitational effects are disregarded). Consequently, any proposed "solutions" offered to the CCP are inevitably segregated into several pairwise incompatible categories that differ in their assumptions about what the future nature of the problem will be.

The immediate goal of the present paper is to explain how it is that such incompatible theoretical proposals can all simultaneously count as plausible solutions to the CCP, when the CCP is itself understood as referring to multiple, mutually exclusive problems that only arise in mutually exclusive possible future contexts. As will be argued in section 4, these mutually exclusive "versions of the CCP" contour a (perhaps surprising) amount of structure within quantum gravity research, suggesting a richer understanding of theory development than that which is conveyed by metaphors of principled explorations of some 'theory space'. Generalizing from the case of the CCP, I will argue that the primary function of certain 'problems', at least in theoretical physics, is to sketch out new avenues of theoretical research. Under this view, what constitutes a possible 'solution' to such a problem is any proposal for the next generation of theory that is well-suited to reproduce the virtues of the present theory and improve upon it in at least one precise regard: by providing an explanation for how it is that the given problem (newly articulable as a unique problem) is already handled in the new theory.

2 The "Cosmological Constant Problem" Is Not a Problem (For Current Theory)

This section ought to begin with a concise statement of the CCP, but to avoid introducing any basic misunderstandings, its statement first requires sufficient familiarity with the physical theories it (allegedly) implicates. Consequently, the section will instead begin with some preliminary details about the standard model of cosmology and the theory of GR upon which the standard model is built.

In the context of GR, one may understand a cosmological model to consist in three objects: a manifold M, a Lorentzian metric g_{ab} defined on M, and a tensor field T_{ab} on M, each of which satisfy certain technical conditions that will not be important here. As a pair, (M, g_{ab}) describes a general relativistic spacetime (c.f. Malament (2012)); meanwhile, T_{ab} encodes the stress-energy of a collection of matter fields over that spacetime. As was said above, the standard model of cosmology is built on the framework of GR, which means that the geometry of the spacetime wholly determines the distribution of matter throughout spacetime (as represented by T_{ab}) via the Einstein Field Equation (EFE). What is the EFE is a matter of minor controversy, but start by considering it in the following form (in geometrized units where $c = G = 1$):

$$R_{ab} - \frac{1}{2}Rg_{ab} = 8\pi T_{ab} \tag{1}$$

where the left-hand side describes the spacetime geometry (i.e. the curvature of spacetime) via quantities determined by the metric g_{ab}, and the right-hand side defines the distribution of the stress-energy of matter across the spacetime. The local

conservation of energy-momentum is satisfied by the purely geometrical fact that the covariant divergence of the left-hand side vanishes.

The source of controversy concerning the EFE arises in the following way. It is not difficult to show that the left-hand side of the EFE can be modified to include an additional "cosmological" term $\overset{\circ}{\Lambda}\, g_{ab}$ without losing the conservation law over T_{ab}, as long as $\overset{\circ}{\Lambda}$ is a fixed constant:

$$R_{ab} - \frac{1}{2}Rg_{ab} + \overset{\circ}{\Lambda}\, g_{ab} = 8\pi T_{ab} \tag{2}$$

In this version of the EFE, the cosmological term is naively interpretable as the inherent elasticity of the spacetime. Historically, when Einstein first presented his own model of cosmology, he used equation (2), but he did not need to. One can easily modify equation (2) to return to an expression like that of equation (1), where the cosmological term $\overset{\circ}{\Lambda}\, g_{ab}$ is absorbed into the stress-energy tensor that governs the right-hand side of the expression:

$$R_{ab} - \frac{1}{2}Rg_{ab} = 8\pi(T_{ab} - \frac{\overset{\circ}{\Lambda}}{8\pi}g_{ab}) = 8\pi T_{ab}^{(total)} \tag{3}$$

The two versions of the EFE are thus equivalent, but in the original version we can understand the collection of matter fields represented by the stress-energy tensor $T_{ab}^{(total)}$ to also include the possibility of a vacuum term: a constant stress-energy defined everywhere on the spacetime.[1] Even when all of the ordinary matter contributions represented by T_{ab} go to zero (i.e. in vacuum regions), there is a constant quantity of

[1] Curiel (2016) offers an argument based on a novel uniqueness proof for the EFE that the "cosmological" term ought only to be understood in this latter context (returning to the original version of the EFE). Although relevant to the subject matter of the present paper, Curiel's argument should not affect the particular claims about methodology and theory development that I will make. For this reason, I leave off a study of its potential implications on desired solutions to the CCP for another time.

stress-energy that characterizes the region. In summary then, a positive-valued "cosmological" term $\overset{\circ}{\Lambda} \, g_{ab}$ is understood as corresponding to a negative-valued vacuum term given by the expression $-\frac{\overset{\circ}{\Lambda}}{8\pi} g_{ab}$.

Whether one understands the EFE as the expression given in equation (2) or as the expression given in equation (3) is a matter of free choice about what one considers to be the theory of GR upon which the standard model of cosmology is built. Since the 1990s, the standard model of cosmology has included a non-zero effective term like $\overset{\circ}{\Lambda}$, here denoted Λ, to recover empirical observations about the accelerating expansion of the universe.

That the standard model of cosmology requires a non-zero Λ is normally taken as evidence that there is such a thing as an 'effective' vacuum term, like that included in equation (3) but precisely characterized by Λ. Moreover, since the vacuum term resembles a constant multiple of the spacetime metric, it is easy to interpret the effective term as arising due to a classically available and gravitating energy density of the vacuum, with a pressure of opposite sign and zero shear (c.f. Zel'dovich (1968), Frieman et al. (2008)). This sets the stage for theoretical considerations about the phenomenology of matter fields that could, in principle, be thought to give rise to such a 'vacuum energy'.

At this point, it is crucial to note that nothing in the theory of GR and, as a consequence, nothing in the standard model of cosmology prohibits hypotheses concerning a new classical term Λ_0 which behaves like a contribution to the effective vacuum term in the form $T_{ab}^{(0)} = -\frac{\Lambda_0}{8\pi} g_{ab}$. This term is often referred to as the "bare cosmological constant" or, in context, the "bare term" (c.f. Carroll (2001); Martin (2012)). Note that, by the same arguments as given above concerning the "cosmological"

term, Λ_0 may be placed on either side of the EFE. In its ordinary usage, for instance as the quantity appears in the Lagrangian/effective field theory formalism popular in contemporary high energy physics, Λ_0 is regarded as a 'zeroth order' term in the gravitational sector of GR, which is just to say for present purposes that, unlike the rest of the contributions to the effective vacuum term that are ordinarily interpreted as *due to matter*, Λ_0 is ordinarily thought to be grouped together with the various curvature terms determined by the metric. Hence, it is natural to think of a non-vanishing Λ_0 as introducing a "cosmological" term on the left-hand side of the EFE as in equation 2, even under the interpretive convention that puts all other contributions to the effective vacuum term together on the right-hand side. In this way, the value of Λ_0 in the standard model of cosmology can be understood, on the standard view, to modify which geometries one properly identifies as 'vacuum regions' according to the underlying theory of GR.

Why one would entertain such a term is secondary to the main point at hand, which is that the standard model of cosmology is built in such a way that it already includes the term, which could either represent a new constant in the underlying theory (i.e. the standard view), or else it could represent a new kind of (classical) matter field in our universe that resembles the other contributions to the effective vacuum term. No matter its interpretation, its function in the model is that of a fit parameter to help match the model to empirical data (even in the case where $\Lambda_0 = 0$). The value given to Λ_0 is wholly determined by the following expression:

$$T_{ab}^{(0)} = -\frac{\Lambda_0}{8\pi}g_{ab} = T_{ab}^{(\Lambda)} - T_{ab}^{(\gamma+\delta+\zeta+...+\theta)} = -\frac{\Lambda}{8\pi}g_{ab} - T_{ab}^{(\gamma+\delta+\zeta+...+\theta)} \tag{4}$$

where $T_{ab}^{(\Lambda)}$ is the effective vacuum term and $T_{ab}^{(\gamma+\delta+\zeta+\ldots+\theta)}$ represents any other classically available contributions to that effective term.

While there are no classical sources arising in any domains of physical theory that resemble the contributions $T_{ab}^{(\gamma+\delta+\zeta+\ldots+\theta)}$, there are quantities in the standard model of particle physics that could be candidates for such contributions, if only our theory of gravity (and subsequently our model of cosmology) were built in such a way as to handle quantum phenomena. Those quantities are the expected zero-point energies that emerge from the fields described by the standard model of particle physics (i.e. in the framework of QFT on flat spacetime) when those fields are arranged in their vacuum states.[2] The presumed relationship between these zero-point energies and Λ may be understood as follows. First, interpret the zero-point energies as the energy densities of the various vacua that arise in QFT; there exist arguments that these zero-point energies have well-defined non-zero and non-infinite expectation values that are computable (in principle, and to some approximation) within the framework of particle theory. Moreover, for such quantities to be invariant with respect to the symmetries of the implicit background Minkowski spacetime of standard QFT, the corresponding (expected) stress-energy must resemble a constant multiple of the Minkowski metric (and so, the zero-point energies must come coupled with a quantity of isotropic pressure of exact proportions and with opposite sign, with zero shear). Insofar as the background Minkowski spacetime can be naively understood as an arbitrary spacetime of GR, one has found a stress-energy tensor of just the form one needs to account for Λ.

Nonetheless, it is important to emphasize that these quantities are quantum

---

[2]This is not the place for an extended conversation about zero-point energies. For slightly more, see section 3; otherwise, for more careful treatments of zero-point energies, see Rugh et al. (1999); Rugh and Zinkernagel (2002); Martin (2012); and Kragh (2012).

mechanical expectation values, which is to say that they are not quantities of energy in the classical sense of the term (Saunders, 2002). Any treatment of these expectation values as classical quantities applicable to a general relativistic model of cosmology depends on semiclassical approximations that model quantum mechanical quantities as classically gravitating substances. But lacking a mature theory of quantum gravity which can be shown to reduce to GR in the appropriate limiting cases, it is unclear why the predicted zero-point energy values from QFT on flat (Minkowski) spacetime can be assumed to gravitate like otherwise classical energy contributions in curved spacetime. (As will hopefully become clear in the following section, such an assumption amounts precisely to a belief in advance about what will turn out to be true, in light of the future theory of quantum gravity that we presently lack.) Moreover, even understanding Minkowski spacetime as a 'vacuum spacetime' (the sort of spacetime on top of which quantized matter fields ought presumably to be defined) requires that one commit by stipulation to the belief that GR is a theory in which the bare cosmological constant Λ_0 vanishes. Naturally, this stipulation undermines the extent to which zero-point energies can be expected to explain or predict any observed value of Λ.[3]

---

[3]I have assumed, in parenthetical, that the sort of spacetimes on which one defines matter fields are vacuum spacetimes, where 'vacuum spacetime' gains its meaning via consideration of the EFE in the theory of GR. In the context of GR, only were Λ_0 to vanish does Minkowski spacetime pick out a geometry that is appropriately identified as vacuum. As one reviewer rightfully points out, there is a sense in which, in the context of the so-called "semiclassical EFE", the situation changes: spacetimes on which one may appropriately define a particular species of quantized matter field are exactly the "semiclassical" vacuum spacetimes, where "semiclassical" vacuum spacetimes (with respect to some species of quantized matter field) are defined as those spacetimes for which the zero-point energies of that species of quantized matter vanish (if Λ_0 is assumed to vanish) or otherwise source a particular non-zero "cosmological" term characterized by Λ_0. Hence, in the case of the semiclassical EFE, zero-point energies feature in a new additional consis-

Regardless of these considerations, physicists regularly equivocate between the effective term Λ and the interpretation of any term like $\overset{\circ}{\Lambda}$ as a vacuum term belonging on the right-hand side of the EFE and due primarily to the expectation values of the zero-point energies that arise in QFT.[4] Different heuristics to compute the total value of the collection of zero-point energies suggest that the quantity is between 60 orders of magnitude (Frieman et al., 2008) and 120 orders of magnitude (Weinberg, 1989) larger than Λ. If this is understood as a rough but direct prediction of Λ, it truly is the biggest (finite) disparity between prediction and observation in the history of physics. When the collection of zero-point energies is instead considered as the "matter contribution" to Λ, then (by the outline sketched above) Λ_0 assumes the difference between the empirically determined Λ and its total computed "matter contributions".

Finally, I am in the position to concisely characterize the CCP according to its three common variations, without fear of perpetuating any odd understandings of the terms involved. The "old" CCP suggests that there is a problem in cosmology that Λ in the

tency constraint concerning the identification of vacuum states in the theory defined on Minkowski spacetime (if Λ_0 vanishes) or another spacetime of constant scalar curvature (if Λ_0 does not vanish).

I have some reservations about using the semiclassical EFE in this way (and any use of the semiclassical EFE certainly takes us beyond our well established theories), but putting those aside, the issue seems to me now to be even more stark: why should quantities derived in QFT formulated on Minkowski spacetime (for instance) say anything at all about what happens in theories formulated on other spacetimes, for which plausible vacuum states (supposing they exist) are subject to entirely different consistency constraints?

[4]This equivocation is not entirely unmotivated; the hope, at least for some, is that all sources of classical spacetime curvature according to the EFE might ultimately be explained by quantum phenomena in the framework of QFT. But I would like to caution against such optimism. It is far from obvious, for instance, why one should expect that the resources available in the present standard model of particles exhaust all possible quantum phenomena that could come to explain Λ in future theories of quantum gravity.

standard model is effectively 0, despite the seemingly overwhelming presence of "matter contributions" that arise as consequences of the standard model of particle physics. The "new" CCP suggests that the problem is that Λ is not precisely 0, while also being many orders of magnitude smaller than the total computed value of those "matter contributions". The "cosmic coincidence problem" suggests that the "new" CCP is made more peculiar than it may otherwise have been by the fact that the particular value of the not-quite-zero Λ resembles the (classical) mass-density of the universe in the present cosmic epoch.

But we have already seen that none of these variations of the CCP constitute strict problems within the standard model of cosmology, because the standard model permits the inclusion of a new term Λ_0, whose assumed value provides, by construction, the difference between the effective vacuum term and whatever other classical vacuum contributions are identified. Moreover, according to current theory, the "matter contributions" are zero-point energies that emerge in the context of QFT, which are precisely not the sorts of classical quantities that readily act as contributions to Λ.

Nonetheless, in framing the subject in these intertheoretic terms, one begins to get the sense that the CCP is foremost an issue of how it is that the consequences of other theories that are, at face value, incompatible with the theory at hand can be leveraged in the physical interpretations of particular terms within the theory at hand. If Λ is interpreted as a vacuum term (i.e. the total gravitating energy of the vacuum), and if the quantum "matter contributions" to Λ arising in QFT on flat spacetime are thought to be exhaustive of all suitable (i.e. gravitating) vacuum physics, then considering the remaining contribution to Λ (i.e. Λ_0) as anything besides a vacuum term seems wrong-footed. On the other hand, if one simply wishes to assign a physical interpretation

to the remaining contribution Λ_0 as a vacuum term, such as by postulating the presence of a constant, classical field, then one is wielding the standard model of cosmology including the physical interpretation of Λ as a criticism of the completeness of another theory entirely. In other words, while the CCP does not describe a problem within our current theories as such, it does suggest problems that may soon arise, as one begins to speculate about the contours of a future theory of physics that is meant to straddle the overlapping domains of the current ones.

Pausing, for just a moment, our discussion of the particulars of the CCP, I suspect it will be helpful to meditate on one of the general themes of this section.[5] The point has been made that the CCP is about a putative conflict between an empirically loaded quantity, i.e. Λ from the standard model of cosmology, and a theoretical entity arising in our premier theory of matter, i.e. zero-point energies. Recalling Laudan (1978), this situation has all the hallmarks of an empirical problem ("...anything about the natural world that strikes us as odd, or otherwise in need of explanation..." [p. 15]), perhaps best regarded as an unsolved empirical problem, given that it has indisputably directed future inquiry.

On the other hand, I have stressed repeatedly that there is no sense in which the theoretical entity arising in our premier theory of matter *decidedly matters* in conversations about Λ in standard model cosmology, and the freedom of the bare term Λ_0 in current theory insulates our current standard model of cosmology from any similar problem in principle. Any conflict here must therefore be understood in exclusively intertheoretic terms; this is the conclusion that we have just arrived at. For Laudan

---

[5]I thank an anonymous reviewer, as well as an editor, for encouraging me to make these general morals more explicit.

(1978), this would seem to place the CCP as a conceptual problem— an external one, in fact. It is true that Laudan regards empirical and conceptual problems as co-existing along a spectrum, or a "continuous shading" [p. 48], of problem types intermediate between two extremes— my intention is therefore not to disparage his taxonomy. To the contrary, what I find so curious about the CCP, and this will become the primary focus in the following section, is that despite being, in some sense, a conceptual problem that straddles our current theories of matter and gravity, it nonetheless seems *most aptly described* as a first-order, empirical problem whose domain or context of inquiry just so happens to be located within some future theory that we simply do not yet have. Since we do not yet have that theory, there will be genuine expert peer disagreement about the precise nature of the problem, which will subsequently motivate radically different paths of research; this is the subject I will take up extensively in the following section. For now though, the slogan to repeat is this: the CCP is *about* future physical theory, identified in advance via the details that we today suspect will come to feature in that future theory.

3   Making Sense of The Problem(s)

It might be tempting to dismiss the CCP outright, and declare the physics community's reactions to it philosophically lazy, misguided, or even agenda-driven. This is the view implied by Earman (2001, p. 207), who writes: "Steven Weinberg (1989), who believes that physics thrives on crises, has been instrumental in promoting this problem to the status of a crisis for contemporary physics. I want to explain why this 'crisis' needs to be viewed with some skepticism."

I do not dispute Earman's claims in these remarks, but I confess that Earman's

general attitude here strikes me as backward. The philosophical point to be made is not that the physics community might be in error about a concern that is central to their field (i.e. that "this 'crisis' needs to be viewed with some skepticism"). Rather, the point to be made is that the physics community's persistent worry about the CCP *as a problem* (alternatively, as a "crisis") suggests that there might be something else going on. There is something interesting, that is, about the fact that the depiction of the CCP as a problem for contemporary cosmology has not diminished over the past several decades in light of all of the potential "solutions" that have been offered for it.

There is an explanation for this state of affairs, which was hinted at in the final remarks of the previous section. If the zero-point energies that arise in QFT are understood as predictions today about what vacuum energy sources will be present in some future physical theory which unifies GR and QFT (i.e. the theory of quantum gravity), and if, moreover, vacuum energies in that future theory contribute to some effective term that reduces to a classical vacuum term, then that future theory (but *not* any present theory) is thought to have to reckon with the apparent disparity between the quantum contributions in current QFT and Λ in the standard model of cosmology. Subsequently, solving the CCP consists in endorsing a possible framework for future physical theory and subsequently demonstrating that within such a framework, the CCP is an ordinary empirical problem with an ordinary solution already built into the future theory. Insofar as any such solution depends on conjectures about what future theory will entail, the CCP has remained (and will continue to remain) an open problem until some of those conjectures gain sufficient empirical grounding to warrant adoption into the corpus of established physical theory. But this entire conversation presupposes some resolution to a fascinating methodological conundrum: how do we talk about what

future, as-of-yet undiscovered theories *ought to entail*, and on what grounds do we make such claims?

In the case of quantum gravity, in which the goal is to find a theory that formally interprets both the gravitational properties of energy sources and the physics associated with quantum states, the standard move is to introduce a semiclassical theory of gravity that is meant, in principle, to approximate such an unknown future theory. Unfortunately, since the eventual theory of quantum gravity is unknown, it is unclear in what respects and what domains of applicability this procedure is meant to approximate that eventual theory, and a well-evidenced classical model of cosmology (e.g. the standard model of cosmology) does not necessarily transform into a well-evidenced semiclassical model when the underlying theoretical frameworks are swapped. Nonetheless, this is the sort of architecture available to theorists interested in considering possible eventual consequences of quantum gravity in cosmological contexts.

The point here is that claims made on the basis of semiclassical variants of classical gravitational models are not consequences of current physical theory. Rather, they are claims about what we might infer from contemporary physics about approximations of future physics. This answers the first part of the conundrum voiced above (how do we talk about what the future theories ought to entail: by semiclassical alternatives to the current theories), but it does not answer the second part of the conundrum (what are the grounds for these claims about such future theories). Notice that justifications for the methods used to predict consequences of future theories can only ever consist in claims which derive from the current theories, which (by setup) are silent precisely in those contexts where the semiclassical theory is needed. Thus, the assumptions that go into these methods are open to expert disagreement: there are reasonable disputes about

what semiclassical gravity entails for the future of cosmology, which means that there can be no obvious answer to the second part of the conundrum. This conclusion is not surprising. Nonetheless, I believe it is helpful to think about theory development explicitly in these terms, i.e. as cases of genuine disagreement about the details of the future theory sought, stemming from disagreements about what it is appropriate to broadly infer about that future theory from our current best theories that we have developed so far.

In fact, competing assumptions along these lines give rise to three broad categories of solutions to the CCP:

1 Assume that vacuum quantities (i.e. quantities that resemble constant multiples of the metric) do not gravitate as ordinary sources in the EFE. Then the CCP becomes: what gives rise to the effective vacuum term characterized by Λ in the standard model of cosmology?

2 Assume that zero-point energies gravitate as tensorial quantities on the right-hand side of the EFE and exhaustively source the effective vacuum term characterized by Λ. Then the CCP becomes: how does one account for the discrepancy between the currently computed values of the zero-point energies and the observed Λ?

3 Assume that Λ is not exhaustively sourced by vacuum energies. Then the CCP becomes: what other physical mechanisms can contribute to that which is presently understood as the effective vacuum term characterized by Λ?

Note that both the second and third categories expect (or at least permit) zero-point energies to contribute to the effective vacuum term that arises in the standard model of

cosmology. (Where they disagree is about what the status of the bare term Λ_0 will be after all vacuum sources have been considered.) In this way, they are both incompatible with category **1** solutions, which assume that the vacuum term represented by Λ in the EFE simply cannot consist in vacuum stress-energies. To get a sense of what this means, consider some examples of category **1** solutions: unimodular gravity (c.f. Earman (2003) and Ellis et al. (2011)) and gravitational aether (c.f. Afshordi (2008)). Both approaches consider trace-free restrictions of the right-hand side of the EFE, and hence vacuum terms are decoupled from spacetime curvature.[6]

By contrast, both categories **2** and **3** assume that vacuum sources, where they can be assigned suitable stress-energies, gravitate in the usual way. But whereas category **2** accepts the semiclassical assumption that zero-point energies in QFT will come to (fully) source Λ, category **3** disputes it. More formally, category **2** assumes the most common semiclassical variant of the standard model of cosmology, in which the term that reduces to Λ is considered to be explained by those terms that reduce to the zero-point energies in QFT. Given this as a starting point, approaches in category **2** focus on undermining

---

[6] The inference here in the case of unimodular gravity has been criticized by Earman (ript) (see, in particular, §4.1 therein). It is perhaps more appropriate to say: assuming first that any vacuum terms are taken to be fixed externally, once and for all, across all physical possibilities intended to be circumscribed by the models of the resulting theory, unimodular gravity decouples those terms from spacetime curvature. Whether or not the antecedent assumption is viable as an attitude in theorizing about cosmology is beyond our purposes here, but will surely inform the sense in which unimodular gravity constitutes a satisfying category **1** solution to the CCP. Per arguments in the same paper, it may also be the case that unimodular gravity, understood as a different theory whose interpretation precludes the assumption just articulated, winds up as a solution to the CCP belonging in a different category— such difficulties seem to have less to do with how the categories are mutually contrasted and more to do with the nuances inherent in delineating theories in terms other than their interpretations (and particularly when those interpreted theories have not been extensively studied).

the plausibility of the particular values handed over from QFT as contributions to Λ. Approaches in category **3**, meanwhile, begin with the assumption that there are other mechanisms that contribute (perhaps exhaustively, perhaps not) to Λ that cannot be understood as stress-energies of vacuum sources. Subsequently, work done in category **3** is spent seeking out alternative physical explanations for Λ's particular non-zero value. Hopefully it is clear how approaches in these two categories, insofar as they are taken to be solutions to the CCP, are thus incommensurable from the start (even though work in one category can absolutely inform work in the other category).

Theorists working in category **2** have two avenues available to them that would solve their variant of the CCP. Both avenues consist in challenging the particular computations of zero-point energies in contemporary particle physics as providing an accurate heuristic for the values of the same in the future theory. The first avenue usually involves disputing the assumption that calculations of vacuum quantities in QFT formulated on flat spacetime resemble plausible predictions of vacuum quantities in QFT formulated on a more generalized family of curved spacetimes. Instead, heuristic accounts of QFT on curved spacetime are explored to determine whether the computed zero-point energies are sufficiently suppressed so that the CCP does not appear. In other words, this avenue is motivated by the possibility that the generalization of the standard model of particle physics to particle physics in the presence of spacetime curvature, on route toward a theory of quantum gravity, will happen to undercut the CCP.[7]

---

[7]For example, when Wald (1994) introduces his candidate for a semiclassical stress-energy tensor that is defined with respect to the fundamental observables of a quantum field theory understood in a setting appropriate for curved spacetimes, the uniqueness of that candidate depends on putting in by hand its value when the relevant quantum field is defined on Minkowski spacetime and arranged in its vacuum state. In other words, it is essential in his approach that one stipulate what is the value of the zero-point energies that

The other avenue consists in suggesting potential modifications to standard particle physics on flat spacetime, even before considering how the model could be generalized to curved spacetimes. Work along this avenue explores what sorts of modifications would, in effect, cancel out the currently computed values. The idea here is that the standard model of particle physics is incomplete, even in the low-energy regime of approximately flat spacetime. Improving upon the model, even still in the framework of QFT on flat spacetime, might happen to resolve the discrepancy between Λ and the total contributions from the zero-point energies. For instance, it was first noted by Zumino (1975) that certain supersymmetric particle theories could suppress zero-point energies in precisely the right way to remove the worry of the CCP. Generalizations of this idea have become quite popular in the context of supersymmetric string theories (see, e.g., Kachru et al. (2003)).

There are many theoretical difficulties with this broad approach, especially if Λ is non-zero (Witten, 2001), but its pursuit as a possible solution to the CCP is noteworthy. The standard model of particle physics is incredibly well supported by empirical data, so the idea is not that the standard model *ought to be* modified so as to address the CCP. Rather, the idea is that there is value in determining which sorts of modifications to the standard model of particle physics happen to solve the CCP. In the case of such modifications, if in the future it were to turn out that such modifications to the standard

are thought to arise in standard model particle physics. Depending on one's arguments about how to generalize QFT to curved spacetimes, one may find themselves stipulating different values, including those that would mimic the effects of the effective vacuum term in the standard model of cosmology. On the other hand, as one reviewer points out, the inflexible nature of this framework as a means by which to resolve the CCP can easily lead to over-corrections, and therefore to a new instantiation of the CCP, as which seems to occur in an article by Hollands and Wald (2004).

model were justified, then the community would no longer be worried about the CCP.

Along both of the theoretical avenues, notice how the CCP, not quite a problem within its present context, is transformed into an ordinary, defeasible problem in the contexts of new theoretical initiatives (e.g. QFT on curved spacetime, supersymmetry).[8] I can be more explicit. A solution to the CCP by appeal to some new theoretical initiative X can be understood as consisting in a demonstration that the CCP would be a solved problem in the next generation of physical theory, were the next generation of physical theory to include X. In category **2** (pursued by those who believe that zero-point energies ought to wholly account for the presence of Λ in the standard model of cosmology), X is any generalization of the standard model of particle physics, which is presently formulated in QFT on flat spacetime, to a new theoretical framework compatible with our present understanding of spacetime qua gravity. In such a generalization, the CCP is transformed into a particular ordinary problem within that framework: what is the total magnitude of the expected quantities of zero-point energies that arise in that framework?

This suggests that one role of the CCP in its present form is to provide a heuristic by which new theoretical initiatives are judged, or rather, explicit motivation for the pursuit

---

[8]It is worth flagging here what I intend throughout this paper by the term "theory", as opposed to "theoretical initiative" or "framework". It would take us too far from the central point here to discuss what ultimately ought to delineate these various constructs, properly construed. As such, except where the word appears as parceled together in a standard phrase from physics (e.g. quantum field *theory*, string *theory*, etc.), I have tried to reserve "theory" for use just in regards to our current, well-supported theories of GR and QFT (formulated on Minkowski spacetime), as well as in regards to the future theory of quantum gravity we ultimately seek. Theoretical initiatives and frameworks, by contrast, are more preliminary in nature, as the sort of things upon which or with the help of one might hope to eventually develop successful cosmological models.

and development of some such initiatives over others. Once suitably transformed, the defeasibility of the CCP (as it is rendered in the terms available to some new theoretical initiative) is viewed as a strength of that theoretical initiative: because were it the case that future physicists turn out to need the new theoretical initiative, a (future) problem has both already been formulated and been subsequently solved. This could partially explain the intense spotlight that has been shined on the CCP over the past several decades: for as long as theoretical physicists have been seriously pursuing a new theory of quantum gravity (and the standard model of cosmology was sufficiently mature so as to warrant talk of Λ), the CCP has provided a heuristic by which to evaluate new proposals, orienting focus in an otherwise wild theoretical field. Amongst those who think that the new theory ought eventually to respect zero-point energies as the exhaustive sources of the vacuum term that arises in present cosmology, solutions to the CCP in category **2** highlight which theoretical initiatives along the way are most attractive.

Contrast the transformation of the CCP relevant to category **2** with that which occurs in the case of category **3**. Recall that category **3** is characterized by its rejection of the common semiclassical assumption that zero-point energies fully source the vacuum term characterized by Λ in standard cosmology. For this category, X is any semiclassical theory of gravity in which the geodesic structure of a spacetime is no longer thought to uniquely determine the distribution of stress-energy across it. Beginning with the assumption that this is to be the next new theoretical initiative, the CCP is transformed into a different ordinary problem with a different ordinary solution: what new physics can at least partially account for Λ, which is presently understood as wholly characterizing a gravitating vacuum term? Approaches in this category generally consider the addition of new fields on spacetime that do not couple to spacetime

geometry as traditional energy sources, or else more fundamental modifications to how we think of the relationship between spacetime geometry and matter.

The first of these two avenues takes seriously the popular interpretation that Λ is, at least partially, due to a "dark energy" that couples in atypical ways with the other matter fields involved in standard particle physics. In other words, one expects there to be an effective classical field theory that can characterize the accelerating expansion of the universe. If it turns out to be a constant scalar field, then it is empirically indistinguishable from the interpretation of the effective vacuum term as due to vacuum energy contributions in the classical regime, but should be empirically distinguishable from predictions of the standard model of particle physics in a high-energy regime described by quantum field theories. If it turns out to be any other sort of field (e.g. the model of quintessence given by Zlatev et al. (1999), or else that of an exotic fluid as given by Kamenshchik et al. (2001)), then it is empirically distinguishable from the current theories in both regimes. Either situation carries implications for an eventual theory of quantum gravity.

The second avenue explores what it would take to capture the accelerating expansion of the universe (whether constant or ultimately dynamic) as a cosmic-scale consequence of certain local geometrical properties of spacetime, independent of the matter content of the universe. In these approaches, the quantity Λ that arises in the current standard model of cosmology is understood as an effective term that captures the magnitude by which GR is inadequate as a framework on which to build cosmic-scale spacetime models. $f(R)$ theories of gravity are phenomenological examples of this approach, but there are others examples as well. For instance, MacDowell-Mansouri gravity (Wise, 2010) and the related projects in "doubly special relativity" or "de Sitter relativity"

consider spacetime theories like GR, except where the most relevant tangent space at each point is de Sitter-like, instead of Minkowskian (Kowalski-Glikman and Nowak, 2003; Aldrovandi and Pereira, 2009; Almeida et al., 2012).

Another example along this second avenue (but of a very different character to those already mentioned) is found in cascading gravity models, which realize *degravitation*, the theoretical notion that the presence of additional large spacetime dimensions can effectively degravitate vacuum energies (De Rham et al., 2008). These approaches take seriously the question "why does the vacuum energy gravitate so little?" (Dvali et al., 2007, p. 1) and try to answer it by demonstrating feasible mechanisms (i.e. different embeddings of four-dimensional spacetime in a higher-dimensional space) by which Λ can be decoupled from any computed values of vacuum energies. The goal of these projects is to dilute the relationship between the observed value of Λ and any computed stress-energies of vacua in precisely the right way to reconcile observations and/or predictions of the two in conjunction.

In all of these cases, note that the modifications to GR obviously influence considerations about semiclassical gravity by providing a new classical theory that semiclassical gravity must reduce to. This in turn guides which strategies toward a theory of quantum gravity are considered most viable. In this way, considerations of the many different sorts of solutions to the CCP that appear in category **3** reveal a richer understanding of the role that the CCP plays in theoretical research: more than providing a guiding hand as to which theoretical initiatives can do more than others in the wild field of quantum gravity research, the CCP suggests possible physical constraints on future theories of quantum gravity that are both motivated by our most well-evidenced theories and that would otherwise be entirely unforeseen. Just like in the

case of the category **2** solutions to the CCP, both avenues of research in category **3** pave the way for the eventual theory of quantum gravity to have a built-in solution to a technical problem that formally reduces to the CCP. But now I can say a bit more: category **2** solutions focus on pushing the boundaries of what we know about the quantum nature of matter based on appeal to cosmological evidence, while category **3** solutions focus on pushing the boundaries of what we know about the relationship between matter and spacetime geometry (i.e. geodesic structure) based on appeal to cosmological evidence. Naturally, each of these is relevant in pursuit of quantum gravity.[9]

4   Discussion and Concluding Remarks

The goal of this paper was to discuss the odd state of affairs surrounding the CCP, in the hopes of drawing philosophical lessons about scientific methodology on the cutting-edge of theory development. To that end, some care was given to how the CCP was introduced: as a quirk of our standard model of cosmology but not technically a problem within it, which only starts to look more like a problem when one turns one's attention to what future physical theories might entail. How the CCP transforms into an ordinary problem in light of what future theories are thought to entail became the focus of section 3, whereupon it was noticed that the CCP transforms into different problems

---

[9]There is a fourth category of solutions to the CCP in the context of future theory that focuses on explaining away the CCP by heuristic arguments concerning probability measures that emerge in the context of several speculative theoretical initiatives, which render the observed value of Λ antecedently probable. Since the backbone of the fourth category, anthropic reasoning, is a source of its own philosophical battles, the details of solutions in this category have not been included in the present work. For more on the subject, see Smeenk (2013) and Curiel (2015) to get a sense of the arguments that plague the general approach.

for future physical theory depending on the assumptions one is prepared to make about what the future theory will look like.

This unifying feature of all three categories– that the CCP of today will eventually take the form of an ordinary problem with an explicit solution already provided– suggests a general understanding of why it is that the CCP of today is treated by the theoretical physics community as a problem at the forefront of their field. Simply put, it is the reduction to the language of our present theories of (one of) the first empirical problem(s) that the next physical theory (but *not* any current theories) will be expected to solve, above and beyond all of the other problems that it must also already solve (namely, all those that reduce to already-solved problems in our current best theories). Moreover, it provides a way by which the consequences of our best current theories can be leveraged for a glimpse at what may be necessary in the development of the next generation of physical theories. That the solutions to the CCP— those glimpses at what may be necessary— can sometimes seem entirely incommensurable as solutions to the same underlying problem turns out to be an artifact of the quite reasonable observation that there exists genuine expert disagreement about what we know today concerning the future of physics.

In regards to the first of these points, Weatherall (2011) identifies a similar, although retrospective, story concerning the equivalence (according to Newtonian physics) of inertial and gravitational mass, where it was viewed as a problem that the two Newtonian quantities were empirically indistinguishable (even though it technically was not a problem for the theory). The solution (that is, an explanation of their observed equivalence) came in the context of GR, even though there is no equivalent problem in the language of GR to be solved (because there are no such things as inertial or

gravitational mass). Nonetheless, in the reduction of GR back to Newtonian physics, one may derive the equivalence of the two Newtonian quantities, which seems to count as a satisfying explanation in response to the original Newtonian problem. In this way, the development of GR is retrospectively characterized as having made progress in regards to the purported problem, even though there was no actual problem in the framework of Newtonian physics and the theory of GR is such that there could never be any such problem, coherently stated.

Weatherall (2011) uses this case to come to the conclusion that at least some problems in physics have the effect of shaping the next generation of research in the field, because their solutions involve departing from present theory and engaging in new theoretical developments. Neglecting the details of his particular argument, I am inclined to make a related but slightly stronger claim on the basis of the CCP: the primary function of at least some problems in physics is to shape the next generation of research in the field, because for researchers to merely entertain them *as problems to be solved* requires new theoretical initiatives with which to properly articulate them. The distinction between Weatherall's claim and my own is subtle but important: my claim is not that problems like the CCP are wanting for solutions in the next generation of theory, whereupon today's theory will be viewed as inadequate in regards to solving such a problem. To the contrary, recall the slogan from the end of section 2: the CCP is *about* future physical theory. That is to say, the broad strokes which characterize research in frontier physics are laid out in accordance with the different ways in which intertheoretic worries like the CCP can be converted into specific articulable problems to be solved, whereupon by the time the details of the next generation of theory are worked out, those brand new specific problems will already have been taken care of.

I contend that the intense focus on the CCP in contemporary theoretical physics supports this claim. As a consequence of the community worrying about what can only really be described (at worst) in present theory as the setting of a fit parameter according to empirical data, multiple independent lines of advanced theoretical inquiry have been developed, each of which features a sketch of future physical theory in which a newly articulated problem has been solved. Moreover, those sketches can be easily grouped, broadly, in terms of the additional, often pairwise incompatible assumptions that would warrant them as components of the next generation of physical theory.

It would be a mistake to miss what is at stake in this claim. That we have access to what are essentially first-order empirical problems within a future theory, without having prior access to the details of that future theory, is extraordinarily counter-intuitive. That potential solutions may be worked out for those problems in advance, prior to writing down that future theory, is perhaps even more so. But I hope that the case study of the CCP demonstrates how these affairs can nonetheless be so, albeit at the cost of there being genuine peer disagreement about what we know today on the basis of our current best theories of physics.

If the target of my hope here has been achieved, I believe that the present analysis runs counter to an otherwise dominating metaphor in the methodology of theoretical research. The dominating metaphor to which I refer is that which considers theory development in terms of principled explorations of something like a 'theory space'. As an instance of this metaphor in action, consider the recent essay by Rovelli (2018), in which he laments, in the context of frontier physics and quantum gravity research, what he identifies as "the current "why not?" ideology: any new idea deserves to be studied, just because it has not yet been falsified; any idea is equally probable, because a step further

ahead on the knowledge trail there may be a Kuhnian discontinuity that was not predictable on the basis of past knowledge..." [p. 487]. The influence of the 'theory space' metaphor is, I think, quite obvious: there exists some total collection of ideas from which the next theory is to be selected, according to some or other principle of selection. Rovelli's charge against the frontier physics research community is that the relevant principle of selection employed is something like random sampling with respect to a principle of indifference. Against this principle of selection, Rovelli goes on to write that "When we consider ourselves to be "speculating widely", we are mostly playing out rearrangements of old tunes: true novelty that works is not something we can just find by guesswork". In other words, the criticism of the "why not?" ideology is (at least in part) one of internal inconsistency: that the sampling mechanism utilized in its name does not conform to the specifications that would otherwise justify it. But Rovelli also points out that the "why not?" ideology suffers, moreover, from that total collection of ideas being too large to meaningfully sample.

Rovelli's conclusion is that we ought instead seek a method that has science proceeding "through continuity, not discontinuity". My implicit suggestion in this paper has been that there are such other methods that may already be found in practice, and that the role of the CCP in quantum gravity research witnesses one of them. Namely, I have suggested that the persistent identification of the CCP as a problem in need of solution exhibits a method of theory development in the frontiers of physics, whereby future theory is built in the first place as the sort of thing which already solves what may today be recognized as empirical problems arising within it. Note the departure from the 'theory space' metaphor: my claim is not that theory development be understood, in this case, in terms of the selection of a new idea from the space that happens to solve the

CCP; to the contrary, the new idea is built in the first place by supposing future theory to already be constrained in certain ways (by the successes of current theory), and subsequently solving the problems that would arise within the future theory given those constraints.

Taking a step back, it is clear that by virtue of being entertained as empirical problems in the context of some or other future theory, intertheoretic worries like that of the CCP in our current best theories provide a way to leverage those current theories as justifications for particular new theoretical initiatives in the pursuit of future physics. That is to say, the CCP is best understood as a means by which the frontier physics research community may illuminate the possible paths forward from current theories of spacetime and matter to future, more sophisticated ones.

References

Afshordi, N. (2008). Gravitational aether and the thermodynamic solution to the cosmological constant problem. *arXiv preprint arXiv:0807.2639*.

Aldrovandi, R. and J. Pereira (2009). de sitter special relativity: effects on cosmology. *Gravitation and Cosmology 15*(4), 287–294.

Almeida, J. B., C. Mayor, and J. Pereira (2012). De sitter relativity: a natural scenario for an evolving λ. *Gravitation and Cosmology 18*(3), 181–187.

Carroll, S. M. (2001). The cosmological constant. *Living Rev. Rel 4*(1), 41.

Curiel, E. (2015). Measure, topology and probabilistic reasoning in cosmology. *arXiv preprint arXiv:1509.01878*.

Curiel, E. (2016). A simple proof of the uniqueness of the einstein field equation in all dimensions. *arXiv preprint arXiv:1601.03032*.

De Rham, C., S. Hofmann, J. Khoury, and A. J. Tolley (2008). Cascading gravity and degravitation. *Journal of Cosmology and Astroparticle Physics 2008*(02), 011.

Dvali, G., S. Hofmann, and J. Khoury (2007). Degravitation of the cosmological constant and graviton width. *Physical Review D 76*(8), 084006.

Earman, J. (2001). Lambda: The constant that refuses to die. *Archive for History of Exact Sciences 55*(3), 189–220.

Earman, J. (2003). The cosmological constant, the fate of the universe, unimodular

gravity, and all that. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 34*(4), 559–577.

Earman, J. (manuscript). Trace-free gravitational theory (aka unimodular gravity) for philosophers.

Ellis, G. F., H. Van Elst, J. Murugan, and J.-P. Uzan (2011). On the trace-free einstein equations as a viable alternative to general relativity. *Classical and Quantum Gravity 28*(22), 225007.

Frieman, J. A., M. S. Turner, and D. Huterer (2008). Dark energy and the accelerating universe. *Annual Review of Astronomy and Astrophysics 46*, 385–432.

Hollands, S. and R. M. Wald (2004). Essay: Quantum field theory is not merely quantum mechanics applied to low energy effective degrees of freedom. *General Relativity and Gravitation 36*(12), 2595–2603.

Kachru, S., R. Kallosh, A. Linde, and S. P. Trivedi (2003). De sitter vacua in string theory. *Physical Review D 68*(4), 046005.

Kamenshchik, A., U. Moschella, and V. Pasquier (2001). An alternative to quintessence. *Physics Letters B 511*(2), 265–268.

Kowalski-Glikman, J. and S. Nowak (2003). Doubly special relativity and de sitter space. *Classical and Quantum Gravity 20*(22), 4799.

Kragh, H. (2012). Preludes to dark energy: zero-point energy and vacuum speculations. *Archive for history of exact sciences 66*(3), 199–240.

Laudan, L. (1978). *Progress and its problems: Towards a theory of scientific growth*, Volume 282. University of California Press.

Malament, D. B. (2012). *Topics in the foundations of general relativity and Newtonian gravitation theory.* University of Chicago Press.

Martin, J. (2012). Everything you always wanted to know about the cosmological constant problem (but were afraid to ask). *Comptes Rendus Physique 13*(6), 566–665.

Rovelli, C. (2018). Physics needs philosophy. philosophy needs physics. *Foundations of Physics 48*(5), 481–491.

Rugh, S. E. and H. Zinkernagel (2002). The quantum vacuum and the cosmological constant problem. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics 33*(4), 663–705.

Rugh, S. E., H. Zinkernagel, and T. Y. Cao (1999). The casimir effect and the interpretation of the vacuum. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 30*(1), 111–139.

Saunders, S. (2002). Is the zero-point energy real? *Ontological aspects of quantum field theory*, 313–343.

Smeenk, C. (2013). Philosophy of cosmology. *The Oxford Handbook of Philosophy of Physics.*

Wald, R. M. (1994). *Quantum field theory in curved spacetime and black hole thermodynamics.* University of Chicago Press.

Weatherall, J. O. (2011). On (some) explanations in physics. *Philosophy of Science 78*(3), 421–447.

Weinberg, S. (1989). The cosmological constant problem. *Reviews of Modern Physics 61*(1), 1.

Wise, D. K. (2010). Macdowell–mansouri gravity and cartan geometry. *Classical and Quantum Gravity 27*(15), 155010.

Witten, E. (2001). The cosmological constant from the viewpoint of string theory. In *Sources and detection of dark matter and dark energy in the universe*, pp. 27–36. Springer.

Zel'dovich, Y. B. (1968). The cosmological constant and the theory of elementary particles. *Soviet Physics Uspekhi 11*(3), 381–393.

Zlatev, I., L. Wang, and P. J. Steinhardt (1999). Quintessence, cosmic coincidence, and the cosmological constant. *Physical Review Letters 82*(5), 896.

Zumino, B. (1975). Supersymmetry and the vacuum. *Nuclear Physics B 89*(3), 535–546.