Problems with the cosmological constant problem

Adam Koberinski* †

Abstract

The cosmological constant problem is widely viewed as an important barrier and hint to merging quantum field theory and general relativity. It is a barrier insofar as it remains unsolved, and a solution may hint at a fuller theory of quantum gravity. I critically examine the arguments used to pose the cosmological constant problem, and find many of the steps poorly justified. In particular, there is little reason to accept an absolute zero point energy scale in quantum field theory, and standard calculations are badly divergent. It is also unclear exactly how a semiclassical treatment of gravity would include a vacuum energy contribution to the total stress-energy. Large classes of solution strategies are also found to be conceptually wanting. I conclude that one should not accept the cosmological constant problem as standardly stated.

1 Introduction: What's the problem?

There is a near consensus among theoretical physicists working on a quantum theory of gravity—particularly those trained in particle physics—that the cosmological constant poses a major theoretical problem.[1] That's where the consensus ends; exactly *what* the problem is

---

*Department of Philosophy and Rotman Institute of Philosophy, University of Western Ontario, London, ON N6A 5BY, Canada

†akoberin@uwo.ca

[1]There is not *complete* consensus, however. See Bianchi and Rovelli (2010) for one notable criticism of the idea that the cosmological constant poses any problem at all.

and *how* it affects our understanding of physics is unclear. In his seminal review, Weinberg (1989) introduces the cosmological constant problem (CCP) as one of the few crises available to serve as fodder for theoretical physicists:

> Perhaps it is for want of other crises to worry about that interest is increasingly centered on one veritable crisis: theoretical expectations for the cosmological constant exceed observational limits by some 120 orders of magnitude. (p. 1)

Others share the worry, but drastically decrease the magnitude of the discrepancy. Burgess (2013) claims that the CCP stems from attempting to answer a simple question: "What is the vacuum's energy, and how does it gravitate?"

> [I]t is a problem because observations indicate a vacuum energy that is more that 50 orders of magnitude smaller than we think it should be, given the particles and interactions that we find in experiments. (p.1)

Others take a different approach to stating the problem. Rather than focus on the magnitude of the discrepancy between prediction and observation, one may wonder why the observed cosmological constant is so small, why it isn't exactly zero, or why the vacuum energy density is on the same order of magnitude as the current matter energy density of the universe (see Nobbenhuis 2006, Section 1).[2] Still others think that all or these are ill-stated, and don't get at the heart of the CCP:

> In some respects, the biggest problem with the cosmological constant problem is that it is rarely stated properly. Why is the cosmological constant so small? Why is it not Planckian? Why is it not zero? All of these questions belittle

---

[2]A brief comment on the nature of the citations given here. I focus on review articles and lecture notes from physicists. Lecture notes are particularly illuminating, since they are used to convince the "uninitiated" graduate students that the problem really is a problem. Given the near-universal acceptance of the CCP among working physicists, these arguments are often unstated in published articles. The review articles are an exception—they often feature critical analyses of proposed solutions, as well as somewhat detailed discussion of the nature of the CCP itself.

the cosmological constant problem, and if we are to have any hope of solving it, we had better be clear about what the problem really is...[T]he real issue with the cosmological constant is not so much one of fine tuning, but of radiative instability, and the need to *repeatedly* fine tune whenever higher loop corrections are included. (Padilla 2015, p.2)

Radiative instability is uniquely a problem for perturbative formulations of the CCP, order-by-order fine-tuning only occurs in a perturbative treatment. A small minority don't think that there even is such a thing as the CCP! Bianchi and Rovelli (2010) are highly critical of stating the CCP as a crises for current physics:

There is no great mystery, in an accelerated expansion of the universe. This is a phenomenon which is clearly predicted and simply described by well-understood current physical theory. It is well understood in the context of general relativity, which naturally includes a cosmological constant...the common theoretical objections against this interpretation of the acceleration are either weak, or ill-founded. (p. 1)

So, with the disagreement amongst physicists about what the problem is, or whether there is a problem at all, it is no surprise that the CCP remains a confusing ground for philosophers. Saunders (2002) has examined the philosophical issues concerning the reality of the quantum vacuum, in analogy with the electromagnetic ether and the Dirac negative energy sea. In the latter cases, physics managed to do away with these concepts while maintaining phenomena they were originally thought to explain. The technical issues with the vacuum in quantum field theory are a bit tougher, however. Rugh and Zinkernagel (2002) provide the first detailed philosophical analysis of the CCP as a whole, and critically highlight some of the conceptual barriers I discuss below. Where they remain largely agnostic regarding the resolution of these barriers, I argue that they provide good reason to reject the CCP.

3

In Section 1.1, I will provide the relevant background to the CCP, and attempt to disambiguate the seemingly distinct issues that get lumped together under its umbrella. I take all of these to contribute to *the* CCP, but each is worth mentioning as a separate piece. Section 2 outlines the background assumptions that one must accept in order to accept the CCP as a real problem. In subsections I detail the steps, and highlight philosophical issues with each. The steps build upon one another; if one rejects any step—as I will argue one should—then the CCP doesn't arise at all. Section 3 puts these issues aside, and outlines some of the broad strategies for solving some form of the CCP, given that one takes it to be a problem in the first place. Though interesting models have been proposed in response to the CCP, none adequately deal with the problem. Coupling the lack of solution with the shaky motivation, I conclude that one should not take the CCP seriously.

1.1 Multiple CCPs

From the previous section, we learn that physicists often speak of different—though related—problems using the phrase "the cosmological constant problem." In fact, introductions to the problem will often list three distinct problems, commonly referred to as the old CCP, the new CCP, and the coincidence problem. The first two are more appropriately thought of as challenges for future models of quantum gravity, rather than issues with the current Standard Model of particle physics of the ΛCDM model of cosmology. Our best theory of matter—quantum field theory—tells us that the vacuum has an enormous energy density. Our best theory of gravity—general relativity—tells us that this energy density should gravitate, and a semiclassical merger of the two indicates that vacuum energy density should couple to the Einstein field equations in the form of a cosmological constant term. What we get from quantum field theory is an enormous cosmological constant term, well beyond what

4

would be physically allowable based on observation.[3] Requiring that the quantum vacuum be Lorentz invariant heavily constrains the form of the vacuum stress-energy tensor. The most general tensor quantity that is Lorentz invariant is a multiple of the Minkowski metric, so $\langle T^{\mu\nu} \rangle \propto \eta^{\mu\nu}$. Through the standard process of minimal gravitational coupling, we replace the Minkowski metric with a general Einstein metric, and introduce the expectation value of the vacuum energy as the correct constant of proportionality,

$$\langle T_{vac}^{\mu\nu} \rangle = -\langle \rho_{vac} \rangle g^{\mu\nu}. \tag{1.1.1}$$

Conservation of stress-energy forces $\langle \rho_{vac} \rangle$ to be constant.[4] Equation 1.1.1 looks like a special case of the stress energy of a perfect fluid, where $p_{vac} = -\langle \rho_{vac} \rangle$, indicating that the equation of state for the vacuum is parameterized by $w = -1$. If we insert $\langle T_{vac}^{\mu\nu} \rangle$ into the Einstein field equations, it takes on a form identical to that of the cosmological constant, with $\Lambda_{vac} = 8\pi G \langle \rho_{vac} \rangle$. By taking the vacuum energy seriously, we seem to "predict" the presence of a cosmological constant in a semiclassical treatment of gravity.

The old CCP assumes that $\Lambda = 0$, and leads to attempts to find mechanisms going beyond quantum field theory and general relativity for eliminating the vacuum energy contribution. However, observations have since indicated that the expansion of the universe is accelerating, meaning that our best cosmological models now include a nonzero Λ. Thus, the new CCP is a bit more challenging: one must find a way of canceling the vacuum contributions, while also including a small, positive effective Λ. Cancellations in light of the new CCP need to be finely tuned to at least 50 orders of magnitude in order to leave a residual $\Lambda \neq 0$, which

---

[3]The discrepancy is so large that the very fact of our existence would preclude a cosmological constant term anywhere near that predicted. Weinberg (1989) famously "predicted" the now observed value of the cosmological constant by placing anthropic bounds on its value.

[4]While $\langle \rho_{vac} \rangle$ must be constant, we should expect ρ_{vac} to vary heavily. The quantum vacuum fluctuates, and these fluctuations are supposed to provide some empirical evidence for the reality of vacuum energy. A fully quantum treatment would force ρ_{vac} to be far from constant.

physicists find unnatural.[5]

Both the old and new CCPs are made more severe by the fact of radiative instability. Unlike certain other quantities in quantum field theory—such as the mass and charge of the electron in quantum electrodynamics—the diverging value of the cosmological constant is not properly renormalizable. Renormalization involves "redefining" constants in a theory in such a way that divergences cancel out. A renormalizable term is cured of divergences at a finite order within a perturbative expansion, and remains cured at all higher orders. The vacuum energy in quantum field theory is radiatively unstable, which leads to the traditional forms of the CCP. Radiatively unstable terms can be tuned to a given order in a perturbative expansion, but they diverge again when higher order terms are included. If the vacuum energy were renormalizable, there would be no CCP at all. One would simply redefine the vacuum energy as the physically observed vacuum energy, and this would be no more puzzling than the mass or charge of the electron. One may be unsatisfied with the renormalization procedure, but this is a separate issue from the CCP. Though not explicitly stated in many formulations of the CCP, the radiative instability of the vacuum energy is extremely important for understanding problems with fine-tuning solutions.

Given the radiative instability of the vacuum energy, a finite value is obtained by placing an explicit cutoff in the integrals of the theory. Though this step is often taken in quantum field theory, renormalizable terms end up without a dependence on the exact value of the cutoff used. The vacuum energy density, by contrast, scales at high energies like σ^4, where σ is the cutoff scale. Though the explicit cutoff value is somewhat arbitrary, some value must be chosen to get a quantitative prediction for the vacuum energy density.[6] The fact that the problem persists across any physically reasonable choice of cutoff is used as justification

---

[5]It's not obvious why a high degree of cancellation resulting in $\Lambda = 0$ is any more natural than one resulting in $\Lambda \neq 0$, but theorists intuitions seem to make the new CCP even more pressing.

[6]The size of the discrepancy in the CCP depend on the exact cutoff used. Weinberg gets a 120 order of magnitude disagreement by taking the cutoff up to the Planck scale, while taking the cutoff to just above energies accessible at the LHC gives something like a disagreement of 50 orders of magnitude.

for the robustness of the problem: even though we don't know the correct cutoff scale for Standard Model quantum field theory, the CCP will arise at any value of the cutoff.

The coincidence problem, which is often grouped together with the CCP, is a rather distinct problem, independent of considerations from quantum field theory. In cosmological models the total stress energy tensor is dominated by various contributions at different time scales. In the early universe, radiation dominates. But because there is a finite amount of radiation, and it's energy density in a region of radius r scales as r^{-4}, the radiation dilutes quickly, and the universe moves into a matter dominated stage. The matter dominated stage is one in which structures can form, and beings like us can evolve. Matter energy density dilutes as r^{-3} and so this epoch is also temporary. Finally, the energy density from the cosmological constant is (unsurprisingly) constant, and the final epoch is dominated by Λ. This is an era of mostly empty space, rapidly expanding and diluting the matter and radiation present in the universe. Based on observation and our best cosmological models, matter contributions (including dark matter) make up just over one quarter of the energy in the universe, while dark energy (in the form of a cosmological constant) makes up the remainder. Thus, it appears we are in an era where the matter and vacuum energy contributions are of the same order of magnitude. If we were to pick a time interval at random from the life of the universe, this would seem an improbable era in which to exist.[7] Even factoring in that we must live in an era with sufficient matter, it seems odd then that we are in a transition era between matter dominated and dark energy dominated. A convincing solution to this coincidence problem would make it less surprising that we find ourselves in this current era, and not some other. This is a mostly unrelated problem, but it does involve the observed magnitude of the cosmological constant, so it often falls under the umbrella of the CCP. As

---

[7]There is a rich philosophical literature on issues like this, and whether likelihoods are being appropriately applied within cosmology. Norton (2010), White (2000), and McCoy (2017) are all rather critical of the use of probabilities or likelihoods in these cosmological settings, while Kotzen (2012) and Benétreau-Dupin (2015) defend their standard use, with slight modifications.

such, I will put the coincidence problem aside for the purposes of this paper.

2   The steps required to accept a CCP

In order to accept the CCP as a problem, one must accept a series of implicit steps. These are rarely mentioned in presentations of the problem, and when they are, it is only in passing.[8] In the three of the four steps outlined below, there are convincing reasons for rejecting the orthodoxy and therefore rejecting the CCP. Regardless of the purpose that the CCP serves in constraining future theories of quantum gravity, it is ill-motivated from the point of view of quantum field theory, and questionable when quantum field theory is stitched together with general relativity. This section will proceed as follows. I will list a step necessary to arrive at the CCP, present a standard argument for it, and raise some objections to its acceptance. Since the steps build off one-another, the objections are independent. That is, if you reject my objections to step one, then continue on to step two. If you accept the objections, then that suffices to reject the CCP in full.

2.1   *Is the vacuum energy real?*

In quantum field theory on Minkowski spacetime, a particular dynamical model is described by an action. For simplicity, I will work with a free scalar field in an external potential. Nothing regarding a vacuum energy term depends on this choice, except where noted. The

---

[8]For example, Padilla (2015) highlights the Lamb shift and Casimir effect as evidence in favour of accepting step 1 (below). In a footnote, he notes:

> However, as Jaffe (2005) has emphasized, the calculation of both of these effects correspond to Feynman diagrams with external legs of Standard Model fields. A true vacuum bubble has no external legs at all, and from a gravitational perspective the diagrams we are interested in only have external graviton legs. (p. 4)

This footnote acknowledges that there is controversy over accepting the CCP, but is not mentioned again in the rest of the lecture notes.

action for a free scalar field ϕ in external potential V is given by

$$S = \int \mathrm{d}^4 x \left[V(\phi) + \frac{1}{2} (\partial \phi)^2 \right]. \tag{2.1.1}$$

The energy density of the field is given by the Hamiltonian density operator

$$\mathcal{H} = V(\phi) + \frac{1}{2}(\partial_t \phi)^2 + \frac{1}{2} \sum_i (\partial_i \phi)^2,$$

where the index i runs over the spatial coordinates. This Hamiltonian density is minimized at some constant field configuration ϕ_0 (depending on the particular form of the potential V). The leading quantum corrections to the Hamiltonian density come from a quadratic expansion of the excitations about ϕ_0 in terms of creation and annihilation operators,[9]

$$\hat{\phi}(x) := \phi(x) - \phi_0 = \sum_k \left[u_k(x)\hat{a}_k + u_k^*(x)\hat{a}_k^\dagger \right]. \tag{2.1.2}$$

Putting equation 2.1.2 into the Hamiltonian density, and integrating over the spatial coordinates gives the total energy operator

$$\hat{H} = \left[\frac{1}{2} \sum_k \omega_k + V(\phi_0) \int \mathrm{d}^3 x \right] + \sum_k \omega_k \hat{a}_k^\dagger \hat{a}_k, \tag{2.1.3}$$

where the term in square brackets is the zero point, or vacuum energy associated with $\hat{\phi}$, and $\partial_t u_k(x) = -i\omega_k u_k(x), \omega_k \geq 0$. Note that the vacuum energy as given in equation 2.1.3 is divergent in two ways. First, one sums over ω_k for all k. This is analogous to summing the ground state energy of an infinite collection of harmonic oscillators, indexed by k. Second, as

---

[9]Strictly speaking, we have first put the field theory in a finite volume first. As we take the volume to infinity, the sums below become integrals over continuous k, and technically the operators become unbounded. These issues are glossed over in most of the physics literature, especially in presentations of the CCP, so I neglect them here.

one takes the volume of space to infinity, the integral over space diverges linearly. For most purposes in quantum field theory, this infinite zero point energy is simply ignored. We are often interested in expectation values of excited states, and the relevant energy quantities are *differences* between the excited and ground states. The vacuum energy term is simply discarded, and the energy operator is defined as $\hat{H} := \sum_k \omega_k \hat{a}_k^\dagger \hat{a}_k$.[10]

In terms of path integrals and scattering amplitudes in quantum field theory, vacuum energy terms appear as bubble diagrams in the perturbative expansion. In calculating scattering amplitudes, bubble diagrams appear as a product of all diagrams with external matter lines in the numerator, and so cancel out the analogous bubble diagrams in the denominator. Thus true vacuum energy terms never appear in calculations in quantum field theory.

If we don't ever use it, what grounds do we have for taking the vacuum energy in quantum field theory to be anything more than an inconvenient feature of quantization that we rid ourselves of when calculating? As mentioned above, the Casimir effect and the Lamb shift are often mentioned as effects that force us to take the vacuum energy seriously. I will start with the Casimir effect, as typical presentations appear to depend explicitly on the vacuum state. Casimir (1948) first predicted that closely spaced, uncharged parallel plates would feel a small attractive force, apparently due to fluctuations in the electromagnetic vacuum energy. If one approximates the plates as infinite, impenetrable boundaries, then the spontaneous creation of virtual photons[11] out of the vacuum is constrained between the two plates, but not outside. Since the plates are impenetrable, only photons of appropriate wavelength (i.e., those for which the plate separation is an integer multiple of their wavelength) can be created between the plates. This causes a decrease in energy density in the gap between the plates,

---

[10]In the context of a scalar field, the "subtraction" of vacuum energy can be justified by demanding that the operators in our theory be normal ordered. Roughly, this means that all operators are redefined such that creation operators are always to the right of annihilation operators. However, normal ordering does not work for more realistic gauge theories.

[11]Virtual particles are short lived fluctuations that go undetected by normal means. In addition to their rapid annihilation, virtual particles are not constrained to obey mass-energy relations, and are therefore often called "off mass-shell."

and therefore a decrease in pressure. The plates then move toward each other with a small attractive force. The effect was measured by Sparnaay (1957), seemingly verifying the reality of quantum vacuum fluctuations.

Jaffe (2005) has called into question this orthodoxy, however. He re-derives the Casimir effect in terms of relativistic forces between charges within the plates and current couplings between the two.

> [T]he Casimir effect gives no more (or less) support for the reality of the vacuum energy of fluctuating quantum fields than any other one-loop effect in quantum electrodynamics, like the vacuum polarization contribution to the Lamb shift, for example. The Casimir force can be calculated without reference to vacuum fluctuations, and like all other observable effects in [quantum electrodynamics], it vanishes as the fine structure constant, α, goes to zero. (p. 1)

On Jaffe's construction, the Casimir effect does not depend on vacuum pressure differences, but simply on loop diagram corrections. This is also a more general construction, since different geometric configurations can lead to repulsive forces between conducting plates, and these are not explainable by appeal to the vacuum pressure. It is clear at least that the Casimir effect need not be thought of as a differential vacuum pressure on the parallel plates. Following Jaffe, the Casimir effect should not be treated as any more special than the Lamb shift. Now I will turn to the Lamb shift, and argue that it—along with other one-loop correction effects—should not be taken as evidence for the reality of vacuum energy.

Lamb and Retherford (1947) discovered a small difference in the energy of 2s and 2p orbitals in the hydrogen atom not predicted by the Dirac equation. In quantum electro-dynamics, the effect can be seen in a one loop correction to the interaction between the hydrogen's proton and electron. The one loop correction is often interpreted as a polar-ization of the vacuum in response to the interaction between proton and electron. This

polarization causes an energy shift in the effective potential, proportional to the square amplitude of the electron's wavefuntion at the centre of the atom. Since all electronic orbitals but the s-orbitals vanish at the origin, the shift only affects orbitals with angular momentum $l = 0$, $\Delta E \approx (4m_e)/(3\pi n^3)ln(1/\alpha)\delta_{l,0}$.

Does vacuum polarization force us to take the magnitude of the quantum vacuum seriously? Recall that the magnitude of the vacuum energy is represented pictorially by the bubble diagrams in a perturbative expansion. Bubble diagrams are Feynman diagrams with *no external matter lines*, and are therefore (by definition) undetectable. What I have been referring to as vacuum polarization effects are at best evidence for fluctuations in the vacuum energy in response to ordinary matter and its interactions. At worst, one might caution against reading Feynman diagrams too literally; then the vacuum fluctuations as represented in loop diagrams might be interpreted as simply components of nonlinear interaction effects between ordinary matter. Effects like the Lamb shift and Casimir effect give evidence for nothing more than energy fluctuations within the vacuum. The actual magnitude of energy contained in the vacuum has no evidential support. Even within quantum field theory itself, there is no reason to trust the prediction of expected vacuum energy.

2.2 How much energy is in the vacuum?

Suppose we reject the arguments against the zero point energies presented in the last section. Given that the vacuum contribution to equation 2.1.3 is divergent, how should we determine the magnitude of the vacuum energy density? In order to get a prediction to compare to cosmological observation, we need some way of managing the divergence. Standard renormalization procedures introduce a cutoff scale into the momentum integrals in scattering amplitudes. As long as the divergent quantities (e.g., charge and mass in quantum electrodynamics) scale logarithmically with the cutoff, the divergences can be safely absorbed into other parameters, and the cutoffs can be taken to infinity with the newly redefined physi-

cal masses and charges, for example. This procedure fails for the divergent vacuum energy, however.[12] Vacuum energy corresponding to a field of mass m scales with σ^4, where σ is the cutoff scale:

$$\rho_{vac} \propto \int_0^\sigma \sqrt{k^2 + m^2}\, k^2 dk \approx \sigma^4. \tag{2.2.1}$$

Estimates of the magnitude of vacuum energy density from quantum field theory start by assuming a reasonable cutoff energy scale, up to which we trust the Standard Model. This is typically taken to be the TeV scale, and so "predictions" of the vacuum energy's contribution to the cosmological constant are on the order of at least $(TeV)^4$. It turns out that the exact scale at which the cutoff is introduced is irrelevant—for *any* realistic cutoff scale, $\rho_{vac} \gg \Lambda_{observed}$. The robustness of the problem under changes to the cutoff is supposed to indicate that the problem is somehow independent of the cutoff.

The problem here lies in the fact that any calculation of the vacuum energy cannot be written in a cutoff independent fashion. Given the more modern view of the Standard Model as an effective field theory, strict renormalizability might not necessarily be required to have a well-defined or physically meaningful vacuum energy. At higher energies, nonrenormalizable terms in an effective field theory are suppressed by some characteristic energy factor, or by the introduction of counterterms in the Lagrangian. However, given our ignorance of beyond Standard Model physics, there are no principled grounds for assuming that the divergences of ρ_{vac} will generically by cured in higher energy theories. Any convincing argument supporting suppression will assume a particular class of high energy solution, and this is problematic. The CCP is supposed to be a generic problem, and if a particular class of successor theories is required to even state the problem, one may naturally question its universality, as well as relevance to current well-accepted theories.

---

[12]Though more sophisticated renormalization techniques exist, such as Wilson and Kogut's (1974) lattice renormalization and 'Hooft and Veltman's (1972) dimensional regularization, they do not succeed in renormalizing the vacuum energy.

Further, the effective field theory framework assumes a clean separation of physics at differing energy scales (Weinberg 1997). When physics separates cleanly at different scales, high energy dynamics have little effect of low energy dynamics, typically in the form of fixing certain "constants" in the low energy domain. These can be empirically determined, and then the high energy theory is screened off from the low energy domain of interest. This separation of scales works very well for local matter theories, but some question its applicability in cosmological contexts, particularly those involving a global cosmological constant (Afshordi 2015). As I will discuss below, certain classes of proposed solutions also violate separation of scales, and therefore the motivation for the CCP is undercut in a solution.

Historically, divergent terms in a quantum field theory are not trusted until some mechanism for their suppression or renormalization has been demonstrated. I believe that the vacuum energy should be no different; cutoff-dependent values for ρ_{vac} should not be taken as good quantitative predictions, nor even estimates for the magnitude of a cosmological constant.

2.3   How should the vacuum energy enter Einstein's field equations?

Suppose, again, that we reject the arguments from the previous two subsections, accept the vacuum energy as real, and accept the cutoff dependent value that is calculated from Standard Model quantum field theories. How should we couple the vacuum energy to gravity? As outlined in Section 1.1, global Lorentz invariance and minimal coupling introduce a vacuum stress-energy contribution of the form shown in Equation 1.1.1. In Minkowski spacetime, Lorentz invariance of vacuum observables properly constrains $\langle \rho_{vac} \rangle$ to be constant. I will go into a bit more detail regarding the process of minimal coupling here.[13]

Starting with the Einstein-Hilbert Lagrangian, we add a matter Lagrangian—in particular

---

[13]Thanks to Michael Schneider for helpful discussion regarding the differences in minimal coupling between general relativity and quantum field theory.

the vacuum energy density contribution—and introduce an interaction of the form $\alpha R \rho_{vac}$ to couple the two Lagrangians. Minimal coupling is the requirement that the interaction term vanishes, or that $\alpha = 0$. This is the simplest way to introduce the vacuum energy into the Einstein-Hilbert action. Now, gravity only couples to the vacuum energy via the Lorentz-invariant measure $\sqrt{-g}\mathrm{d}^4x$:

$$S = \int \mathrm{d}^4x\sqrt{-g}\left\{\frac{1}{2\kappa}R + \langle\rho_{vac}\rangle\right\}. \tag{2.3.1}$$

Deriving the equations of motion by varying the action leads to a stress energy tensor of the form $T^{\mu\nu} = -\langle\rho_{vac}\rangle g^{\mu\nu}$. Since minimal coupling works so well in quantum field theory, this appears to be the most natural way to couple the vacuum energy to gravity. Even if more complicated couplings are introduced, if variations of the action are determined perturbatively, this zeroth order minimal coupling term will appear in much the same way.

This raises a few philosophical concerns regarding semiclassical treatments of gravity. We know that classical general relativity is incompatible with quantum field theory, and semiclassical treatments of the two merged can differ depending on the route one takes. As presented above, the priority lies in defining quantities on a highly symmetric Minkowski background, and introducing gravity as a coupling between the metric and Minkowski matter. In effect, this is equivalent to treating spacetime curvature due to gravity as a perturbation to a Minkowski background space. Though certainly the majority view amongst theorists working on a quantum theory of gravity—and the basis for string theory—other approaches to merging general relativity and quantum field theory exist. In particular, some physicists argue that a major desideratum for quantum gravity is that, like general relativity, it should be a background-independent theory.[14] This means that a background Minkowski spacetime is not assumed, and then quantities like the vacuum energy cannot be defined as they currently

---

[14]It's not exactly clear that general relativity is best understood as a background independent theory, but the insight that the dynamics play an important role in dictating the spacetime structure is well-established.

are in ordinary quantum field theory (Smolin 2006). These differences may be reflected in other approaches to semiclassical gravity, where quantum fields are defined on curved spacetime backgrounds. Hollands and Wald (2010) have developed an axiomatic framework for quantum field theory on curved spacetimes, and vacuum energy does not arise in this context, since a vacuum state is not in general well-defined.

> One of the key insights is that—apart from stationary spacetimes or spacetimes with other very special properties—there is no unique, natural notion of a "vacuum state" or of "particles." Indeed, unless the spacetime is asymptotically stationary at early or late times, there will not, in general, even be an asymptotic notion of particle states. (p. 86)

Though decidedly against orthodoxy, this remains a viable avenue of research, and completely does away with the CCP by privileging curved spacetimes over a Minkowski background.

Even in the standard treatments involving minimal coupling, the use of the expectation value of the vacuum energy may be suspect. Wang, Zhu, and Unruh (2017) emphasize that a more fully quantum treatment of the vacuum energy would highlight that it is far from constant.

> Instead, at each spatial point, the spacetime sourced by the vacuum oscillates alternatively between expansion and contraction, and the phases of the oscillations at neighboring points are different. In this manner of vacuum gravitation, although the gravitational effect produced by the vacuum energy is still huge at sufficiently small scales (Planck scale), its effect at macroscopic scales is largely canceled. (pp. 1-2)

If Wang et al. are correct, then the "predicted" cosmological constant is an artifact of an improperly classical treatment of the vacuum energy in quantum field theory. An appro-

priately quantum treatment then leads to exponential suppression of the vacuum energy at macroscopic scales.

Though minimal coupling of $\langle \rho_{vac} \rangle$ to gravity is certainly an open option, it is by no means the only available means for coupling quantum matter to gravity. Given that it leads to such a terrible "prediction" of the cosmological constant, alternative avenues like those just sketched should be more widely pursued. If any alternative to the orthodox coupling turns out to be correct, then the CCP is likely to disappear again.

2.4   Does anything else act like Λ?

If we reject the concerns raised in the previous three steps, we are left with a cosmological constant "prediction" from quantum field theory that is at least 50 orders of magnitude larger than the observed Λ used in the concordance model of cosmology. If the quantum vacuum energy is the lone source of the cosmological constant, then this is a huge problem—even a "crisis." However, there may be other exotic physics that leads to additional counterterms in the Einstein field equations, such that

$$\Lambda_{observed} = 8\pi G \langle \rho_{vac} \rangle + C_1 + C_2 + \cdots , \tag{2.4.1}$$

where each C_i is a contribution from some other physical mechanism. In effect, this would be an attempted solution to the CCP, since the new counterterms would presumably be high energy effects from a theory of quantum gravity. However, it seems reasonable to first treat the sources we know of; I do not see any problems in this final step, and include it only for completeness.

As stated above, if any of the arguments presented in Sections 2.1–2.3 are accepted, then the CCP is no longer a problem. Even if the arguments are rejected, I hope to have shown that there is far more subtlety to the CCP than standard treatments suggest. Setting

aside these concerns, however, in the next section I look at some broad classes of potential solutions.

3 Plausibility of attempted solutions

Because of the appeal of the CCP, there are nearly as many proposed solutions as there are physicists who have heard of the problem. In this section, I will outline a few different strategies for a solution, and some problems that prevent any from standing out as the best way forward. Given the criticisms of the formulation of the problem outlined above, the lack of a clear solution strategy is further reason to reject the CCP as worth pursuing.[15]

3.1 't Hooft naturalness

't Hooft naturalness, also known as technical naturalness, is a mathematical means of shielding otherwise unconstrained parameters from being too large. If a theory contains a parameter that would go to zero with an added symmetry, and that added symmetry is not too badly broken by the original theory, then the scale of symmetry breaking "protects" that parameter from getting too large. This notion of naturalness, as the name implies, is due to 't Hooft (1980). The classic example is the pion mass due to near chiral symmetry. If chiral symmetry—the symmetry between left-handed and right-handed particles—were an exact symmetry in nature, the pseudoscalar meson masses would be forced to zero. Spontaneous breakdown of global chiral symmetry would then lead to the presence of massless Goldstone bosons. Nambu and Jona-Lasinio (1961) originally posited that the pion would be the Goldstone boson related to chiral symmetry breaking in the limit of equal mass fundamental states—what we would now call quarks. Since the quarks are *not* massless nor

---

[15]For more in depth reviews of different approaches to solving the CCP, review articles such as Nobbenhuis (2006) are a good place to start. Here I am concerned simply with common conceptual features of broad classes of solutions, and do not expand on any technical details.

of equal mass, neither is the pion, but the pion's relatively small mass is protected by the approximate flavour symmetry between the up and down quarks. As the bare masses of the up and down quarks approach equality, the pion mass approaches zero. Naturalness thus explains why the pion mass is so much smaller than other quark bound states. The hope regarding the CCP is that one can find an approximate symmetry under which the vacuum energy would be zero, to help explain its small observed value in the form of a cosmological constant.

The most well-known naturalness solution is supersymmetry. Vacuum energy contributions from bosons and fermions are of opposite sign, so if each known particle had a superpartner with the same quantum numbers, all contributions would cancel out, resulting in total vacuum energy of zero. Standard theoretical predictions require $M_{SUSY} \propto \Lambda^{1/2}$, and the lack of observed superpartners up to the currently accessible TeV scale means that supersymmetry—if it exists—is too badly broken to protect the cosmological constant.

Other symmetries have been proposed to protect the vacuum energy, including conformal symmetry. Conformal symmetry forbids any terms in a Lagrangian that set a length scale, so the vacuum energy would be excluded in a conformally invariant theory. Unfortunately for the plausibility of conformal invariance, mass terms are also disallowed. However, conformal invariance can be used as a motivation for modifying the Einstein-Hilbert action, in effect rewriting the action in terms of the trace-free component of the Riemann tensor. A residual cosmological constant term then comes out as a boundary term for the action. This leaves open the question of how the quantum vacuum should factor into gravity.

The problem with most of the approaches utilizing 't Hooft naturalness is that the symmetries would need to be far more exact empirically to dramatically reduce the magnitude of the vacuum energy as standardly predicted. Outside of supersymmetry, naturalness does not address the "prediction" of $\langle \rho_{vac} \rangle$ from quantum field theory. If the CCP is taken seriously

in its standard form, something other than naturalness seems to be needed to solve it.

3.2   Violate equivalence principle

If one believes in the vacuum energy from quantum field theory, a plausible way to dissolve the CCP is to propose a mechanism to decouple the vacuum energy from the Einstein field equations. In effect, this is a violation of the equivalence principle, since then only nonvacuum energy would gravitate. I include in this class of solutions solutions that generalize the equivalence principle as well. In one class of proposals, extra dimensions are introduced, and the large vacuum energy leads to high curvature and compactification of all but the familiar four spacetime dimensions of classical general relativity. In this class of solutions, the equivalence principle is technically not violated, but only appears to be violated due to our ignorance of the extra dimensions. Though compatible with string theories, the addition of new submillimeter scale dimensions is a high price to pay for a solution to an ill-motivated problem. A further mechanism must also be included to explain why all but four dimensions would compactify due to $\langle \rho_{vac} \rangle$, as well as why the equivalence principle —at least approximately—holds for nonvacuum energy in four dimensions.

Further strategies involve infrared modifications to the gravitational interaction, leading to effectively massive gravitons (e.g., Arkani-Hamed et al. (2004)), so-called "fat gravitons" (e.g., Sundrum (2004)), or symmetry breaking where gravitons appear as the Goldstone bosons associated with the broken symmetry (e.g., Weinberg and Witten (1980)). A benefit of infrared modifications is that small corrections to general relativity may be observable at galactic or even solar system scales, making this class of attempted solutions empirically testable. However, many of these solutions end up also violating the assumption that gravitational physics has a clean separation of energy scales. Microscopic phenomena (such as modifications to graviton couplings) end up directly altering cosmic expansion rates. As mentioned in Section 2.2, the cutoff dependence of ρ_{vac} depends conceptually on an effective field

theory framework, in which there is clear separation of energy scales. This brings up a worry that these sorts of modifications undermine the very problem they wish to solve. To this point in time, modifications to general relativity have not been observed, and these models are therefore constrained by observation in the degree to which they can modify gravity at observable scales.

3.3 Statistical solutions

Considerations from quantum cosmology and inflation have led many physicists to take seriously the prospects that our universe is just one in an ensemble of universes, making up the multiverse. It is further conjectured that the fundamental constants of nature—such as the electron charge, strong coupling constant, and the cosmological constant—are variable between universes. The ensemble is assumed to be so vast that all physically possible values of the constants are realized in some universe. We can thus put bounds on possible values of Λ with anthropic considerations (Weinberg 1989). Despite the numerous concerns involved in defining a measure over an ensemble of unobservable universes, this sort of "solution" to the CCP does not explain why the apparently large vacuum energy from quantum field theory does not gravitate in the form of a cosmological constant.

If we look at the wavefunction of the universe, it should in general be a superposition of all physically possible values of the cosmological constant. Hawking (1984) argued that the wavefunction of the universe would be heavily peaked around $\Lambda = 0$. If correct, this might help solve the old CCP, but the new CCP becomes more puzzling in light of Hawking's work. Both the anthropic and quantum statistical solutions to the CCP suffer from the same sorts of defects. First, neither address the vacuum energy or its lack of gravitation. They aim only to make the observed Λ less surprising, and disconnect the quantum vacuum from the cosmological constant. Second, there is no good way to delimit which values of Λ are physically possible, and which aren't. So we have an ill-defined space of possibilities

from which to sample. Even if that problem can be surmounted, there are no grounds for introducing a measure over the possibility space, and thus statistical arguments cannot be justified here. Norton (2010) argues that statistical arguments like these are doomed to fail in cosmology; in any case, they do little to solve the CCP.

4   Conclusions

I have argued against taking the CCP seriously. There are major hurdles to be overcome in (1) believing in an absolute value for the bare vacuum energy, (2) trusting the cutoff dependent perturbative calculation of $\langle \rho_{vac} \rangle$, and (3) inserting the expectation value of the vacuum energy into the Einstein field equations. If one rejects any of these steps, then the CCP—at least as standardly presented—does not arise. Suppose that the CCP is meant to be a guiding problem in constructing a quantum theory of gravity. Then perhaps it needn't be well supported so long as it serves its heuristic purpose. If we look at the sheer number of attempts to solve the CCP, then it has undoubtedly been successful as a heuristic motivator. However, no clear solution strategy has led to major progress in developing quantum gravity. Many of the solution strategies outlined in Section 3 are either ad hoc—their purpose is to solve the CCP, so they have been constructed to do little more than that—or inspired by solution strategies proposed independent of the CCP. Many do not even solve the CCP as traditionally stated. If successful, these "solutions" would have nothing to say about the quantum vacuum.

The issues behind accepting the CCP highlight the reliance on heuristics in arguments in high-energy physics. Due to the lack of empirical access to high-energy phenomena, problems are formulated based on extrapolation of methods and heuristics that have been successful in low-energy particle physics. Often this is not a problem—symmetry arguments, gauge theory, and locality are generally taken to apply at all energy scales—but the heuristics seem

to break down in the case of the CCP.

It is possible that a better argument supporting the CCP can be formulated. If so, it would be interesting to see the details. Unless it can avoid all of the concerns I have raised in Section 2, I am skeptical that it should be accepted. However, even if the problem can be strengthened to avoid some of my concerns, the improved generality may help to better highlight the assumptions involved in the CCP, as well as the best way to "solve" it in a new theory of quantum gravity.

References

Afshordi, Niayesh (2015). "Why is High Energy Physics Lorentz Invariant?" In: *arXiv preprint arXiv:1511.07879*.

Arkani-Hamed, Nima et al. (2004). "Ghost condensation and a consistent infrared modification of gravity". In: *Journal of High Energy Physics* 2004.05, p. 074.

Benétreau-Dupin, Yann (2015). "The Bayesian who knew too much". In: *Synthese* 192.5, pp. 1527–1542.

Bianchi, Eugenio and Carlo Rovelli (2010). "Why all these prejudices against a constant?" In: *arXiv preprint arXiv:1002.3966*.

Burgess, CP (2013). "The cosmological constant problem: Why its hard to get dark energy from micro-physics". In: *arXiv preprint arXiv:1309.4133*.

Casimir, Hendrick BG (1948). "On the attraction between two perfectly conducting plates". In: *Proceedings of the KNAW*. Vol. 51. 7, pp. 793–795.

Hawking, Stephen William (1984). "The cosmological constant is probably zero". In: *Physics Letters B* 134.6, pp. 403–404.

Hollands, Stefan and Robert M Wald (2010). "Axiomatic quantum field theory in curved spacetime". In: *Communications in Mathematical Physics* 293.1, pp. 85–125.

Hooft, G t and M Veltman (1972). "Regularization and renormalization of gauge fields". In: *Nuclear Physics B* 44, pp. 189–213.

Jaffe, RL (2005). "Casimir effect and the quantum vacuum". In: *Physical Review D* 72.2, p. 021301.

Kotzen, Matthew (2012). "Selection biases in likelihood arguments". In: *The British Journal for the Philosophy of Science* 63.4, pp. 825–839.

Lamb, Willis E and Robert C Retherford (1947). "Fine structure of the hydrogen atom by a microwave method". In: *Physical Review* 72.3, p. 241.

McCoy, CD (2017). "The implementation, interpretation, and justification of likelihoods in cosmology". In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics.*

Nambu, Yoichiro and Giovanni Jona-Lasinio (1961). "Dynamical model of elementary particles based on an analogy with superconductivity. I". In: *Physical Review* 122.1, p. 345.

Nobbenhuis, Stefan (2006). "Categorizing different approaches to the cosmological constant problem". In: *Foundations of Physics* 36.5, pp. 613–680.

Norton, John D (2010). "Cosmic confusions: Not supporting versus supporting not". In: *Philosophy of Science* 77.4, pp. 501–523.

Padilla, Antonio (2015). "Lectures on the cosmological constant problem". In: *arXiv preprint arXiv:1502.05296.*

Rugh, Svend E and Henrik Zinkernagel (2002). "The quantum vacuum and the cosmological constant problem". In: *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics* 33.4, pp. 663–705.

Saunders, Simon (2002). "Is the Zero-Point Energy Real?" In: *Ontological aspects of quantum field theory*, pp. 313–343.

Smolin, Lee (2006). "The case for background independence". In: *The structural foundations of quantum gravity*, pp. 196–239.

Sparnaay, MJ (1957). "Attractive forces between flat plates". In: *Nature* 180.4581, pp. 334–335.

Sundrum, Raman (2004). "Fat gravitons, the cosmological constant and submillimeter tests". In: *Physical Review D* 69.4, p. 044014.

't Hooft, Gerard (1980). "Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking". In: *Recent developments in gauge theories. Proceedings of the NATO Advanced Study Institute on recent developments in gauge theories, held in Cargèse, Corsica, August 26-September 8, 1979*, pp. 135–157.

Wang, Qingdi, Zhen Zhu, and William G Unruh (2017). "How the huge energy of quantum vacuum gravitates to drive the slow accelerating expansion of the Universe". In: *Physical Review D* 95.10, p. 103504.

Weinberg, Steven (1989). "The cosmological constant problem". In: *Reviews of modern physics* 61.1, p. 1.

— (1997). "What is quantum field theory, and what did we think it is?" In: *arXiv preprint hep-th/9702027*.

Weinberg, Steven and Edward Witten (1980). "Limits on massless particles". In: *Physics Letters B* 96.1-2, pp. 59–62.

White, Roger (2000). "Fine-Tuning and Multiple Universes". In: *Nous* 34.2, pp. 260–276.

Wilson, Kenneth G and John Kogut (1974). "The renormalization group and the ϵ expansion". In: *Physics Reports* 12.2, pp. 75–199.