The multiple realizability of general relativity in quantum gravity

Rasmus Jaksland*

Forthcoming in *Synthese*†

Abstract

Must a theory of quantum gravity have some truth to it if it can recover general relativity in some limit of the theory? This paper answers this question in the negative by indicating that general relativity is multiply realizable in quantum gravity. The argument is inspired by spacetime functionalism – multiple realizability being a central tenet of functionalism – and proceeds via three case studies: induced gravity, thermodynamic gravity, and entanglement gravity. In these, general relativity in the form of the Einstein field equations can be recovered from elements that are either manifestly multiply realizable or at least of the generic nature that is suggestive of functions.

If general relativity, as argued here, can inherit this multiple realizability, then a theory of quantum gravity can recover general relativity while being completely wrong about the posited microstructure. As a consequence, the recovery of general relativity cannot serve as the ultimate arbiter that decides which theory of quantum gravity that is worthy of pursuit, even though it is of course not irrelevant either qua quantum *gravity*. Thus, the recovery of general relativity in string theory, for instance, does not guarantee that the stringy account of the world is on the right track; despite sentiments to the contrary among string theorists.

Keywords

quantum gravity, general relativity, spacetime functionalism, entanglement, emergent gravity, multiple realizability, string theory

---

*Department of Philosophy and Religious Studies, NTNU – Norwegian University of Science and Technology, Trondheim, Norway. Email: rasmus.jaksland@ntnu.no

1 Introduction

Spacetime functionalism, as advocated by Knox (2013, 2017) with inspiration from Brown (2005), is a prominent recent view of spacetime in general relativity and beyond. As its name suggests, spacetime functionalism identifies spacetime, not with a physical kind, but instead with a function: "the metric field is spacetime because of what it does [...] and not by way of what it is" (Knox, 2017, 3). The present paper finds inspiration in this functionalist conception of spacetime for the conjecture that general relativity as a whole is not linked to a particular physical kind, but rather takes the form of a role that might therefore be realized by multiple types of microstructure. In this light, the place of general relativity in the assessment of theories of quantum gravity is reconsidered. Based on three case studies, the paper finds reasons to expect that general relativity will prove to be multiply realizable among possible theories of quantum gravity; known or unknown. The paper argues on these grounds that we should not be too impressed by the recovery of general relativity in theories of quantum gravity.

This question proves important given the current state of quantum gravity research where any theory is yet to deliver testable empirical consequences. It is to be decided by non-empirical means which theory of quantum gravity that is worthy of pursuit and funding, and here stories about the successes of the theory are centre stage. One such story often alluded to in the string community is the natural occurrence within this theoretical framework of gravity and general relativity:[1]

> String theorists sometimes say that string theory already made at least one successful *prediction*: it predicted gravity! [...] There is a bit of jest in saying so – after all, gravity is the oldest known force in nature. I believe, however, that there is a very substantial point to be made here. String theory is the quantum theory of a relativistic string. In no sense whatsoever is gravity put into string theory by hand. It is a complete surprise that gravity emerges in string theory. Indeed, none of the vibrations of the *classical* relativistic string correspond to the particle of gravity. It is a truly remarkable fact that we find the particle of gravity among the *quantum* vibrations of the relativistic string (Zwiebach, 2009, 11).

> The first general feature of string theory, and perhaps the most important, is that general relativity is naturally incorporated in the theory (Becker et al., 2007, 3).[2]

But how impressed should we actually be by the "remarkable fact" that a theory of quantum gravity features the graviton and can "incorporate" general relativity? These

---

[1]The reader is referred to Dawid (2006, 2013a,b) for a more thorough analysis of non-empirical confirmation in the context of string theory. Consult Dawid (2009) for general epistemic assessment of the status of string theory.

[2]See also Greene (1999, 145) and Polchinski (1998, 5).

and other remarks by proponents of string theory can give the impression that the theory, as a result of these successes, is somehow guaranteed to have some truth to it. In one sense, this holds trivially: if general relativity is recovered in some limit of a theory of quantum gravity, then at least the theory is (apparently) empirically adequate in that regime. However, the confidence in parts of the string community – implicit in the references above – is indicative of the stronger sentiment that the successful recovery of general relativity proves string theory to be on the right track.[3] Obviously, this does not amount to the claim among its proponents that string theory in its current form is exactly true, but neither merely that it promises to be empirically adequate in some limit.[4] String theory provides a novel account of the fundamental constituents of reality by positing a microstructure of two-dimensional relativistic strings (and D-dimensional branes, possibly induced by the strings) from which general relativity is recovered as large scale behaviour. The conviction in parts of the string community seems to be that the successful recovery of general relativity promises that this theoretical framework will not prove to be a dead end in quantum gravity research; that some aspects of the stringy account of the world – perhaps most likely those related to the recovery of general relativity – will find their way to the theory of quantum gravity that will eventually be vindicated.[5] In light of this, the question to be explored here is whether the recovery of general relativity in some limit of a prospective theory of quantum gravity *guarantees* that at least some elements of the theory – in particular the microstructure it posits – will find recognizable counterparts in the actual theory of quantum gravity of our universe. It will here be proposed that this question should be answered in the negative. Instead, it is contended that there are at least indications that the routes to general relativity in quantum gravity are rather generic. It is therefore to be expected that multiple theories of quantum gravity – known or unknown – with very different microstructure can recover general relativity. A theory can therefore recover general relativity while being completely wrong about the microstructure it posits; the recovery of general relativity is not a guarantee that string theory is on the right track. What is the true theory of quantum gravity will, according to this argument, not be decided on the grounds of which theory recovers general relativity.

The argument to this effect has functionalism as its point of departure and is inspired by the observation of Lam and Wüthrich (2018) in the context of spacetime functionalism that "it is seen in general as a virtue of functionalism that it allows – indeed, accounts for – multiple realizability" (5). Since functions are roles rather than physical kinds, they are realized not by what the realizer is, but by what it does. As such, functions are independent of the internal constitution of the realizer; they do not depend on the

---

[3]Indeed, 'on the right track'-talk is a recurring theme among string theorists as documented by Camilleri and Ritson (2015).

[4]Thank you to an anonymous reviewer of *Synthese* for pressing me on this point.

[5]It is charitable to observe that string theory has other features to recommend it such as indications toward the recovery of the Standard Model via D-branes (Antoniadis et al., 2003), providing the expected entropy for black holes (Strominger and Vafa, 1996), and renormalizability. Still, in most accounts the recovery of general relativity is presented as the prominent success of string theory.

nature of the microstructure that constitutes the function in a particular circumstance.[6] If something is a function, then this explains how multiple systems – irrespective of how different their internal constitution might be – can realize it; something just has to play the right functional role. It is particularly this aspect of functions that will inspire and inform the discussion of the recovery of general relativity in quantum gravity in general and string theory in particular. The idea is simply this: If sufficiently large parts of the complex described by general relativity display this insensitivity to the nature of the underlying microstructure, then general relativity might be multiply realizable in quantum gravity. Here the category 'quantum gravity' broadly denotes the possibly vast class of known and unknown theories that – in the terminology of Linnemann and Visser (2018) – can be characterized as *prospective* microtheories: theories that posit a common microstructure from which both quantum mechanics and general relativity might potentially be recovered in such a way that the theory can account for their individual success.[7] Thus, the argument seeks to establish that multiple theories in this class of quantum gravity theories can be expected to recover general relativity while positing very different microstructures.[8] This ambition also explains the restriction of the class of quantum gravity theories to microtheories. While general relativity is trivially recovered in semi-classical gravity – with or without higher order corrections – and various effective field theories that include the Einstein-Hilbert action, such theories do not posit any microstructure and the recovery of general relativity is simply achieved by stipulation. Consequently, they cannot serve as evidence that general relativity can be recovered by multiple types of microstructure. For the purposes of this paper, they are not proper theories of quantum gravity.

The multiple realizability of general relativity directly bears on the significance of the recovery of general relativity by a theory of quantum gravity. If multiple theories of quantum gravity with different types of microstructure can recover general relativity, then a theory can evidently do so and still be completely misguided, i.e. it is possible that none of its elements, the posited microstructure in particular, will find their way to the true theory of quantum gravity of the actual world. The simplest route to this conclusion would be to show that several known prospective theories of quantum gravity do recover general relativity while positing very different microstructure. However, as argued in section 4 neither of these premises are clearly satisfied by the current field of prospective theories of quantum gravity. Instead, the strategy of section 5 is to show how general relativity – or more precisely the Einstein field equations (EFEs) – can be recovered from elements that are either manifestly multiply realizable or at least generic in nature

---

[6]The functionalism assumed here is an ontological functionalism in the terminology of Bealer (1997).

[7]Exactly what qualifies a theory as a prospective microtheory – and thus a potential candidate of a theory that can reconcile general relativity and quantum mechanics – shall not concern us here. The important point is simply that the class of theories of quantum gravity includes not only the actual theory of quantum gravity, but also microtheories of other possible worlds with varying similarity to our own.

[8]Whether general relativity is multiply realizable *within* any one theory of quantum gravity is not at issue here. See Huggett (2017) for a discussion relating to this other theme in the context of string theory.

in the varying circumstances of three case studies: induced gravity, thermodynamic gravity, and entanglement gravity. In all three cases, the recovery of the EFEs does not depend on the exact details of the microstructure realizing it: the recovery only depends on generic elements such as unspecified quantum fields, thermodynamic properties, and entanglement. Each approach indicates that the EFEs can inherit this insensitivity to the nature of the microstructure and with it, the possibility that systems with very different microstructure can realize them; having three different approaches to this effect only bolsters the conclusion.[9] If this holds, i.e. if general relativity is multiply realizable in quantum gravity, then the "remarkable fact that we find the particle of gravity among the *quantum* vibrations of the relativistic string" and that "general relativity is naturally incorporated in the theory" are no guarantee that the stringy account of the world is on the right track! More generally, the present paper argues that there are indications that theories of quantum gravity can recover general relativity without their posited microstructure being even approximately that of the actual world.

One might object that no one would actually claim that the recovery of general relativity implies even approximate truth, and that the conclusion is therefore trivial. I think the attitude in parts of the string community disproves this, but regardless, the present argument serves to inform a related important question also alluded to above: how to prioritize for instance funds and resources between different quantum gravity research programs. The argument of this paper indicates that we might have to reconsider the significance assigned to the recovery of general relativity in our preliminary assessment of theories of quantum gravity. The recovery of general relativity is one of the two primary desiderata for any theory of quantum gravity and as such, it does have some confirmatory significance in quantum gravity.[10] Rather, what this paper claims to inform is how we should update our credence in a prospective theory of quantum gravity upon the discovery that it recovers general relativity. If we use Bayes' theorem to update our credence in a theory of quantum gravity upon the discovery that it recovers general relativity, the prior probability of recovering general relativity in such theories, $P(GR)$, enters in the denominator:

$$P(QG|GR) = \frac{P(GR|QG)P(QG)}{P(GR)} \tag{1}$$

where $P(QG)$ is the prior credence in some theory of quantum gravity, $P(QG|GR)$ is the updated credence upon the discovery that it recovers general relativity, and $P(GR|QG)$ is the probability that it actually recovers general relativity (if the recovery of general relativity is a formal result, then this latter probability is arguably one). All are estimated within the context of the class of theories of quantum gravity. Arguably, this prior probability of recovering general relativity, $P(GR)$, is significantly larger if general

---

[9]Having three different, independent approaches leading to the same conclusion establishes its "robustness" beyond what is achieved by each individual approach (Wimsatt, 1981). However, as observed by Weisberg (2006): "Robustness analysis helps to identify robust theorems, but it does not confirm them" (742).

[10]I would like to thank an anonymous reviewer of *Synthese* for pressing me on this issue.

relativity is multiply realizable, than if it is in a one-to-one relation with a specific (possibly unknown) microstructure. So if we update our credence in light of evidence based on a Bayesian inference, then our credence in a theory of quantum gravity that recovers general relativity, $P(QG|GR)$, is less than otherwise, if general relativity is multiply realizable, i.e. if the argument of this paper goes through. We should, in that case, not be too impressed by the recovery of general relativity in a theory of quantum gravity including string theory.[11] As a result, the argument might adjust the relative weight of this and other factors – UV-completion, universality, structural uniqueness, recovering quantum mechanics, to name some – involved in prioritizing between different quantum gravity research programs[12] in the context of pursuit.[13] In summary, when this paper purports to indicate that general relativity is multiply realizable in quantum gravity, then this is not only meant to (possibly) update individuals' credence in current and future prospective theories of quantum gravity, but also to contribute to the discussion of the appropriate relative volume of research in the competing research programs.

The paper proceeds as follows: Section 2 reviews the recovery of general relativity in string theory and, based on this, suggests a sufficient condition for the recovery of general relativity – that the EFEs can be derived in some limit – which will be operational in the remainder of the paper. Section 3 then details multiple realizability and how the string theory derivation does *not* display this. Section 4 explores, but ultimately rejects, an argument that general relativity is multiply realizable based on its prospective recovery in different current theories of quantum gravity. Instead, section 5 gives an argument to the same effect based on the generic nature of the route to the EFEs in induced gravity (5.1), thermodynamic gravity (5.2), and entanglement gravity (5.3), respectively. The paper concludes that there are indications that general relativity is multiply realizable in quantum gravity.

2   General relativity and string theory

It is helpful to begin the discussion with a brief review of what string theory can actually do when it comes to recovering general relativity. This both serves to set the bar for what has so far been achieved by a theory of quantum gravity and suggests what to

---

[11]To question the significance of the recovery of general relativity in string theory is not in itself novel and has already been argued to be part of a larger debate where "string theorists and their critics typically adopt different attitudes to the heuristic significance of solved, partially solved, and unsolved problems" (Camilleri and Ritson, 2015, 54). However, when it comes to the recovery of general relativity in string theory, the dispute over its heuristic significance has only ensued as the debate whether this problem is solved or only partially solved by string theory (see footnote 16). In contrast, the present argument directly informs the heuristic significance of recovering general relativity in a theory of quantum gravity and contends that it should be moderated, if general relativity, as argued, is multiply realizable in quantum gravity.

[12]See the contributions to Dardashti et al. (2019) for some recent instalments in the debate over the role and relevance of such factors in theory assessment in fundamental physics.

[13]See Cabrera (2018) for a discussion of the difference between the context of pursuit and the context of justification in relation to string theory.

look for in other approaches that promise to deliver general relativity. For a rigorous treatment, see Callan et al. (1985). For a more philosophical exploration, see Huggett and Vistarini (2015).

First, a bit of set-up: consider a closed string, i.e. a string whose ends meet, on a Lorentzian manifold known as target space; due to consistency constraints, target space is 26-dimensional in bosonic string theory and 10-dimensional in superstring theory, but this shall not concern us here. Whereas a particle only extends in one dimension (time), a string extends in one spatial dimension in addition to the time dimension; strings are two-dimensional objects that span worldsheets instead of worldlines. The action for the relativistic point particle is its mass multiplied by the proper time along its worldline, and analogously, the string action is given in terms of the "proper" area swept by the string and the mass per unit length of the string. For the relativistic point particle, the parametrization of the worldline entering the action holds no physical significance, and we therefore require that all physics must be invariant under reparametrization of this parameter. Similarly, the worldsheet is parametrized in terms of two coordinates which hold no physical significance either. String physics must therefore be invariant under a reparametrization of these worldsheet coordinates, but since the action also features the induced metric on the worldsheet that depends on these coordinates, the physics must in addition be Weyl invariant; invariant under change of local scale. Together, these are equivalent to conformal invariance, and any consistent string theory must therefore be conformally invariant. This entails that the field theory defined by the string action is a two-dimensional conformal field theory.

As usual, one can arrive at equations of motion by varying the action. Intuitively, the string equations of motion will admit various vibration modes of the string and indeed, in a particular gauge (a part of) the equations of motion take the form of a wave equation (Zwiebach, 2009, section 7.3). Upon quantizing, these modes are quantum states that can be expressed in terms of creation and annihilation operators acting on a vacuum state. Among the first order excitation modes of the closed, relativistic, bosonic string are the graviton (a massless, symmetric, traceless tensor field with two Lorentz indices) and the dilaton (a massless scalar field). This is the "remarkable fact" that gravity naturally occurs in string theory: "We never put in a dynamical metric [...], yet somehow, the quantum states of the gravitational field have emerged!" (Zwiebach, 2009, 292). From the perspective of the relation between string theory and general relativity, the graviton mode is interesting since it perturbs the metric of the background manifold where the closed string lives. A curved background can thus be conceived as a coherent state of closed string graviton modes. The same goes for the dilaton, whose fluctuations, it turns out, also contribute to the metric perturbation as seen from the point of view of general relativity (known as the Einstein frame).

The derivation of the EFEs begins with a string moving in such a background of a coherent state of closed strings. The derivation assumes a low-energy limit. Since energy – in units $c = \hbar = 1$ – scales like inverse length, this is the limit where the energy is much smaller than the inverse of characteristic length scale of the theory, which in the case of string theory is the string length, l_s. Thus, the low energy limit is equivalent

to the limit where $l_s \to 0$. This entails that massive modes of the closed string can be ignored since they scale as l_s^{-2} (in Minkowski background). In the limit $l_s \to 0$, the massive modes therefore decouple such that only the massless modes – the graviton and the dilaton – contribute.[14] In this limit, one can therefore use an effective action for the two-dimensional worldsheet that takes the form of an interacting two-dimensional field theory that couples to the background metric and the dilaton field.

Using the path integral formalism, the generating functional associated with the string action takes the form of a path integral (functional integral) over worldsheet metrics of manifolds and embedding functions from the worldsheet to target space. The dilaton field can be divided into a fluctuating part and constant part equal to its vacuum expectation value that only depends on the worldsheet topology. Consequently, this constant part of the dilaton field can serve as an expansion parameter for an expansion of the generating functional in topologies; the leading order contribution coming from worldsheets that are topological spheres and first order correction from topological tori. A bit of imagining serves to suggest that worldsheets with spherical topology correspond to tree level and the tori to first loop order; something that is indeed confirmed by the identification of the string coupling, g_s, with the constant part of the dilaton field. Thus, we can take the limit $g_s \to 0$ in addition to the low energy limit $l_s \to 0$. Taking both limits, the generating functional only gets contributions from the metric field and the fluctuating part of the dilaton field and takes the form of a path integral over metrics of spherical worldsheet manifolds.

However, even at tree level, observables (correlator functions) evaluated using this generating functional diverges. To regulate the divergences, a cut-off is introduced, and the fields will generally depend on this cut-off, and this threatens to break the conformal invariance of the world sheet. At tree level, conformal invariance is only restored if the derivative of the background fields with respect to the cut-off vanishes, i.e. if the background fields do not depend on the cut-off anywhere. From perturbation theory in string coupling and string length one can find expressions at tree level – valid in the limit $g_s, l_s \to 0$ – for the derivative with respect to the cut-off of the metric and dilaton fields. Using the constraint that these must vanish everywhere, one finds for a bosonic relativistic string:[15]

$$0 = \frac{D - 26}{6l_s^2}(\nabla\Phi)^2 - \frac{1}{2}\nabla_A\nabla^A\Phi$$
$$0 = R_{AB} + 2\nabla_A\nabla_B\Phi. \tag{2}$$

where Φ is the dilaton field, ∇ is a covariant derivative, R_{AB} is the Ricci tensor, D is the number of dimensions of target space and A and B are Lorentz tensor indices of target space. These two equations are constraints that the background fields must satisfy in order for the physics to be consistent with a relativistic string moving in the

---

[14]Strictly speaking, the massless modes of the bosonic string also includes the Kalb-Ramond field, and this derivation of the EFEs can therefore be regarded as one where the Kalb-Ramond field is assumed to vanish. This, however, makes no difference for the present purposes.

[15]See Callan et al. (1987) for higher order corrections.

background. Interestingly, the constraints reduce to the vacuum EFEs, $R_{AB} = 0$, for constant Φ (since Φ only enters with derivatives). This is the first indication that these consistency constraints recover general relativity.

While the equations (2) were obtained as the renormalization group flow for the regulated world sheet action, they can also be regarded as equations of motion for the background fields through which the string propagates. We may therefore change the perspective and search instead for a target space action that has (2) as its equations of motion. The action arrived at has some similarities with the Einstein-Hilbert action, and upon combining the metric field, g_{AB}, with the fluctuating part of the dilaton field, $\tilde{\Phi}$,

$$\tilde{g}_{AB} = e^{-4\tilde{\Phi}/(D-2)} g_{AB} \tag{3}$$

one arrives at the Einstein-Hilbert action with a matter content comprising a kinetic term coming from the fluctuations of the dilaton field. Finally, the full EFEs are the equations of motion for the Einstein-Hilbert action: "In short, conformal invariance of string theory as a worldsheet QFT entails general relativity for target space" (Huggett and Vistarini, 2015, 1168). In their reflection on the derivation, Huggett and Vistarini (2015) argue that the "result shows that graviton [...] coherent states must be related appropriately: by the EFEs" and observe that "the field equations are simply low energy descriptions of the string itself" (1169). General relativity is recovered from consistency constraints as an approximation from first order perturbation theory in the string length and string coupling for the closed string.[16]

The derivation of the EFEs in string theory is the benchmark for the later exploration of other routes to general relativity in quantum gravity and it can thus serve as a guide for what it means when we say that a theory of quantum gravity recovers general relativity. In light of the string theory derivation, the following condition will be assumed to be *sufficient* for recovering general relativity in quantum gravity:

> **Definition:** A theory of quantum gravity is said to recover general relativity
> if the theory describes a system comprising the whole universe with dynamics
> that in some coarse-grained limit are governed by the (semi-classical) EFEs.

The qualification 'semi-classical' is added to admit that the matter can couple to the metric of the spacetime via the expectation value of the energy-momentum operator.

The qualification that the system comprises the whole universe is included to deal with the worry that the uninterpreted EFEs merely relate components of two tensors

---

[16]Critics argue that this derivation of the EFEs is still short of recovering general relativity, because it – in perturbing around a fixed background – is not properly background independent (e.g. Smolin (2006, 185-186) and Rovelli (2013)). Some string theorists have simply denied this proposing a weaker notion of background independence, while others argue that this is merely an artefact of the background dependent perturbative string theory used for the derivation, but maintain that there are indications that the (unknown) non-perturbative formulation of string theory is background independent (Camilleri and Ritson, 2015). We shall set this issue aside, but observe that the same worry applies to the three routes to the EFEs considered in section 5. They may therefore also be background dependent and thus short of full derivations of general relativity.

to one another. We can describe many different waves – water waves, electromagnetic waves, sound waves etc. – using the wave equation, and so, one might object, could the dynamics of a system be governed by the EFEs, but only feature pseudo-gravity; something analogous to general relativity but qualitatively different such as analogue black holes (Weinfurtner et al., 2013).[17] Such examples should not inform whether general relativity is multiply realizable since realizations, for instance in the form of analogue gravity, are instances where general relativity is also part of the best account of the system realizing this pseudo-gravity. To assess this objection, it is useful to adopt some terminology from the interpretation of dualities: the distinction between an internal and external point of view of a theory. Based on work by Dieks et al. (2015), Sebastian de Haro (2017) defines "[a]n external point of view, in which the meaning of the physical quantities is externally fixed" (116). This is contrasted with the internal point of view which obtains "if the meaning of the symbols is not fixed beforehand" (de Haro, 2017, 116). As de Haro observes, these labels are meant to track the available resources with which to interpret a theory: "'External' suggests that there is a relevant environment to which the theory may be coupled; an external context that fixes the meaning of the symbols. 'Internal' suggests that such context is absent" (de Haro, 2017, 116). If a theory is a theory of the whole world, de Haro argues, then only the internal point of view is possible. In that case, there is no external context with which to fix the meaning of the symbols entering the theory, i.e. to fix what physical quantity each symbol refers to. There is no external point of view since every point of view is encompassed by the theory. All there is to say about the system is given by the internal relations within the theory. Now, if two systems governed by the EFEs are claimed to be qualitatively different in such a way that one features gravity whereas the other only features pseudo-gravity, then this requires an external point of view: the possibility to assess that the symbols of the theory refer to different physical quantities between the two systems. Theories of quantum gravity, however, aspire to be theories of everything.[18] If EFEs govern the dynamics in some – possibly coarse-grained – limit of these theories, then the EFEs must inherit the original aspiration of the theory: to cover all elements of the universe, or at least those that are relevant in the limit covered by the EFEs. In quantum gravity, there is no external point of view, and it shall further be contended that this, consequently, is so for genuine recoveries of general relativity. Thus, general relativity is recovered whenever a theory of quantum gravity – that by definition is a theory of the whole universe – has some coarse-grained limit where the dynamics is governed by the EFEs.

In requiring the system to be *governed by* the EFEs, the sufficient condition seeks to rule out cases where a model of the theory merely chances upon a Lorentzian manifold with matter content that happens to solve the EFEs. Suppose for instance that an empty

---

[17]See Dardashti et al. (2015) and Crowther et al. (2019) for a discussion of the significance of such analogue black holes. See Barceló et al. (2011) for a general review of analogue gravity.

[18]Though we will assume this throughout, more is, as observed by Crowther and Linnemann (2017), to be said about the relation between theories of quantum gravity, fundamentality, and the final theory of everything.

universe model of a theory yields Minkowski spacetime – a solution to the vacuum field equations. This is insufficient to recover general relativity, because the accordance with the EFEs could simply be an accident: there might another (possibly more generic) reason why the empty universe model yields Minkowski spacetime than that the metric curvature relates to the matter by the EFEs. To satisfy the proposed sufficient condition for recovering general relativity, Minkowski spacetime should relate to vacuum *because* this is how curvature relates to matter in the EFEs (without cosmological constant). This qualification is already relevant in string theory, where the background target space is assumed to be some Lorentzian manifold and therefore one that, with the right matter content, might solve the EFEs. However, this would obviously not qualify as recovering general relativity, for the reasons above. It is also important to emphasize that assuming a Lorentzian background does not, by the same argument, amount to simply assuming general relativity. This qualification is important because all the three case studies of section 5 assume Lorentzian background. To satisfy the condition for recovering general relativity, it must be shown that each approach recovers the EFEs as the approximate dynamics in some limit, and thus that the EFEs are the equations of motion for the whole class of models that admit this approximation. In other words, while the case studies assume a Lorentzian manifold in their set-up, this does *not* amount to assuming general relativity. Rather, the point of the case studies is exactly to argue that it follows from this and some additional generic assumption varying among the case studies that the relation between the curvature of the manifold's metric and the matter content is governed by the EFEs. As observed, this mirrors the derivation of the EFEs in string theory: the string theory derivation and the case studies are equal in this regard. The case studies are therefore telling of whether this type of derivation, and thus the recovery of general relativity as we now have it in string theory, can be expected in many prospective theories of quantum gravity.

3 Multiple realizability

Like functionalism in general, multiple realizability is a most common theme in the philosophy of mind: "The multiple realizability thesis about the mental is that a given psychological kind (like pain) can be realized by many distinct physical kinds" (Bickle, 2016). The thesis, as we shall presently conceive of it, is that a mental state – something at the psychological level of description – has multiple realizations at the physical level of description.[19] In this sense, multiple realizability goes against the one-to-one correspondence between mental states and brain states associated with identity theory in the philosophy of mind: "The identity theory of mind holds that states and processes of the mind are identical to states and processes of the brain" (Smart, 2017). According

---

[19]Depending on one's view on reductionism in the context of philosophy of mind, the talk of levels of description might be regarded with suspicion. However, in the present context it seems uncontroversial that general relativity belongs to one length scale and quantum gravity to smaller length scale; at least if one does not promote this to a claim about ontological levels (see Le Bihan (2018) for a discussion).

to identity theory, all creatures – or generally entities – that realize some mental state must have a physical kind in common that this mental state can be identified with.[20]

In its contrast to functionalism, this *identification* is the central thesis of identity theory: whereas functionalism argues that mental states are roles, identity theory argues that they a physical kinds. Similarly, one might imagine a similar debate about whether general relativity is to be identified with a role or a (complex of) physical kinds. With the present focus on multiple realizability, we shall set this semantic/metaphysical question aside, though we might observe (again) that multiple realizability sits well with functionalism. The contrast with identity theory here merely serves to raise the question whether there is a one-to-one or a one-to-many relation between the general relativistic level of description and that of quantum gravity.

In summary, the present paper has no ambition to say what general relativity *is*, but only to assess whether we can expect to recover it in one or many theories of quantum gravity. What bears on this question is whether the recovery of general relativity relies on idiosyncrasies of a particular microstructure and its associated (type of) theory of quantum gravity, or whether general relativity can be recovered from generic elements expected to feature in multiple theories of quantum gravity. The derivation of the EFEs in the context of string theory can serve to illustrate this difference. From the perspective of multiple realizability, we should inquire about the elements entering this derivation and whether they themselves are multiply realizable or at least of a generic nature. We must, in other words, explore whether this derivation is insensitive to the exact nature of the microstructure at the level of description of quantum gravity. In this regard, the string theory derivation of the EFEs rules against multiple realizability. As shown above, string theory offers a route to the EFEs based on the Polyakov action in a curved background and consistency constraints for the worldsheet. But the Polyakov action is specifically the action for a relativistic object extended in time and one spatial dimension – a string – and the consistency constraints are imposed based on reasoning coming from reparametrization of the worldsheet swept by the string. If the string theory derivation were the only route to the EFEs, then it would seem that the EFEs were closely related to a particular sort of microstructure – that of relativistic strings – at the level of description of quantum gravity and thus in a one-to-one rather than one-to-many relation. As stated, the aim of this paper is to show that it might be otherwise: that we have reasons to expect there are many routes to the EFEs – and thus the recovery of general relativity – starting at different microstructures.

4 Towards an argument

As stated, the simplest route towards an argument for multiple realizability is to investigate whether different prospective theories of quantum gravity recover the EFEs in some limit. If they do, then this indicates that they can recover general relativity. If

---

[20]For identity theorists, this entails a identification between the mental state and its realizer, but this semantic component of the functionalism/identity theory debate shall not concern us here.

the theories in addition can be argued to exemplify different types of microstructure at the level of description of quantum gravity, then this would be evidence that general relativity is multiply realizable. At a glance, this approach has some promise. In causal set theory, a Lorentzian manifold with gravitational dynamics is conjectured to obtain as an approximation of discrete causal sets (Dowker, 2006). In loop quantum gravity, something resembling the EFEs can be obtained by first finding a semi-classical approximation and then taking a classical limit (Wüthrich, 2017). In string theory, as shown above, the EFEs (with a particular field content) do obtain as equations of motion at first order perturbation theory in the string length and string coupling. Thus, an optimist might say that general relativity promises to occur in several prospective theories of quantum gravity. Furthermore, causal set theory, loop quantum gravity, and string theory could be argued to exemplify different types of microstructure since they, according to the quantum gravity taxonomy of Huggett and Wüthrich (2013), are exemplars of three different types of theories: causal set theory is a non-metrical lattice theory, loop quantum gravity builds on superposed spin network states, and string theory is a quantum field theory that incorporates gravity.

This approach, however, faces immediate difficulties. First, the recovery of the EFEs is debated for all of the mentioned theories; even string theory.[21] While string theory has produced a rigorous derivation based on consistency requirements of the EFEs, this is in ten (or 26) rather than four dimensions. In covariant loop quantum gravity, the Regge action can be recovered from spinfoam amplitudes under certain additional, i.e. more than consistency, constraints. While the Regge action can be regarded as a type of discretized version of the Einstein-Hilbert action, there is no general mathematical proof that the Regge action converges to the Einstein-Hilbert action in the continuum limit (for more details, see Han (2017) and references therein). Though promising, this is still short of a rigorous derivation of the EFEs in loop quantum gravity. In causal set theory, recovering the EFEs in some limit is still only a hope, and so is a formulation of its quantum dynamics. Also the independence of these prospective theories of quantum gravity has been cast into doubt. This questions the argument that these theories of quantum gravity really exemplify different types of microstructure. There are indications of a deep relation between string theory and loop quantum gravity (Jackson et al., 2015; Mertens et al., 2017), and one can only speculate whether similar relations might obtain between these and causal set theory. The approach towards multiple realizability from the realization in different theories of quantum gravity relies on finding theories different enough to at least render it very unlikely that they share a common microstructure. The more general tenet of this second worry is therefore that the (conjectured) realization of general relativity in some prominent theories of quantum gravity might merely be an artifact of the type of theories we can imagine and find it worthwhile to pursue.

Again, an analogy to the case of consciousness can help exemplify this issue. Suppose you and your friend for some time have had the discussion whether consciousness is multiply realizable. You know that your friend is soon to attend the huge Intergalactic

---

[21]For some recent instalments of this debate, see Lam and Esfeld (2013), Oriti (2014), Carlip (2014), and Huggett (2017).

Gathering for All Creatures (IGAC) and you challenge her therefore to bring home different (apparently) conscious creatures to prove her point that consciousness is multiply realizable. You wait for her return in great anticipation, and with increasing excitement upon her message that she is bringing home three creatures from the IGAC. You imagine they could be weird slimy things, semi-transparent wisps, or green men with antennas. Big is your disappointment, therefore, when she arrives accompanied by a dog, a chimpanzee and your mutual friend Jack who also attended the IGAC. This of course does nothing to persuade you that consciousness is multiply realizable. First of all, that these were a selection of creatures from the IGAC makes no difference once you realize that all of them are from Earth. It seems likely that the dog, the chimpanzee, and Jack are not different enough to indicate multiple realizability and furthermore, while you are rather confident that Jack is conscious, you are less sure about the chimpanzee, and even less so about the dog. Presumably, these are very particular creatures among all the many different creatures attending the IGAC. They were probably chosen by your friend because they were easily recognized as possibly conscious creatures. Many of the other lifeforms present at the IGAC must have been very alien perhaps to the degree that it was difficult to assess whether they were alive at all, let alone conscious. So despite the initial disappointment, you can understand why your friend brought such mundane creatures home from the IGAC: these were the sort of creatures she was looking for. But precisely for this reason, they must be viewed with suspicion if used as a defence of the multiple realizability of consciousness. Being the sort of creatures she was looking for makes it all too likely that they do share a common physical/chemical kind that then is the realizer of consciousness.

This story should serve as a warning when it comes to multiple realizability of general relativity in quantum gravity. For all we know, the class of prospective theories of quantum gravity is vast, and it seems overconfident to presume that causal set theory, loop quantum gravity, and string theory are the slime, wisp, and Martian of quantum gravity rather than the dog, the chimpanzee, and Jack. Just because these are theories out of the vast class of quantum gravity theories, they may well be too similar to tell us anything about multiple realizability; because these are the kind of theories we can imagine to look for. With this possibility, it is unjustified to rely too heavily on the particulars of these most prominent theories in the argument for the multiple realizability of general relativity. Instead, it seems more justifiable to look at general aspects of quantum gravity and whether these can be indicative of an answer to the inquiry. As such, the account finds inspiration in the arguments for emergent gravity reviewed by Linnemann and Visser (2018) that "do not assume any specific QG model, but rather consist of pointing out features of gravity which are (at least typically) characteristic of systems with underlying microstructure" (2). As might be expected, this approach provides more implicit answers to the question of multiple realizability, but arguably it does so on epistemologically safer ground.

5 Three routes to the EFEs

The argument involves three case studies. The first appeals to Sakharov's (1967) induced gravity: that the Einstein-Hilbert action, and thus the EFEs, automatically emerges at first loop order of quantum field theories on Lorentzian manifolds (Visser, 2002). This is used to argue that the EFEs are generic to theories of quantum gravity. The second builds on the discovery by Ted Jacobson (1995) of a relation between thermodynamics and the EFEs.[22] Due to the versatility of thermodynamics, it will be argued that this relation establishes the possibility that general relativity is multiply realizable. Inspired by Jacobson's pioneering work, indications of a close relation between entanglement constraints and the EFEs in quantum gravity have appeared.[23] Arguing that entanglement is a generic and necessary component in any theory of quantum gravity, the relation between entanglement and gravity serves as yet another route to the multiple realizability of general relativity in quantum gravity.

Notice that the argument is *not* simply an inference to multiple realizability from the recovery of general relativity in these three accounts in the fashion of the argument above based on causal set theory, loop quantum gravity, and string theory.[24] Instead, the argument is meant to show that thermodynamic gravity – and similarly for induced gravity and entanglement gravity – is insensitive to the nature of the degrees of freedom realizing it and that general relativity can therefore be recovered as a thermodynamic phenomenon in systems with very different microstructure. Because of the possible thermodynamic nature of general relativity and the multiple realizability of thermodynamics, general relativity is multiply realizable. In this fashion, each account on its own indicates that general relativity is multiply realizable due to its origin in each account in multiply realizable elements or at least elements insensitive to the exact nature of the microstructure: first order quantum corrections in induced gravity, thermodynamics, and entanglement, respectively.

As mentioned in section 2, all three accounts play out on a Lorentzian manifold, and none of them can as a consequence testify whether Lorentzian manifolds are multiply realizable. Indeed, both induced gravity and thermodynamic gravity are grouped by Carlip (2014) as "models in which spacetime, the metric, and diffeomorphism invariance are present, but the dynamics of gravity is emergent" (201); and the same holds for entanglement gravity. As a consequence, they arguably only suffice to support that gravitational dynamics is insensitive to the microstructure realizing it and therefore possibly multiply realizable. For all said here, the assumed Lorentzian manifolds might be in a one-to-one correspondence with a particular microstructure at the level of description of quantum gravity. We will set this worry aside since it applies to the string theory

---

[22]See also Padmanabhan (2010b) and Verlinde (2011).

[23]See for instance Ryu and Takayanagi (2006), Van Raamsdonk (2010), Maldacena and Susskind (2013), Chirco et al. (2014), Lashkari et al. (2014), Jacobson (2016), Han and Hung (2017), Baytaş et al. (2018), Cao and Carroll (2018), and Chirco et al. (2018).

[24]It could not be since none of the three approaches are theories of quantum gravity, but merely intermediate structures from which the EFEs can be derived.

derivation as well as the three case studies below. However, in the broader context of quantum gravity, of course, assuming a Lorentzian manifold is by no means innocent: Recovering a Lorentzian manifold is a most prominent problem in loop quantum gravity (Wüthrich, 2017) and causal set theory (Dowker, 2006) and often equated with the issue of how spacetime emerges in quantum gravity (see for instance Huggett and Wüthrich (2013) and Oriti (2014)). Again, neither of these accounts hold any implications for this question about the emergence of a Lorentzian manifold; though recent research suggests a relation between spacetime and entanglement coming out of entanglement gravity (Van Raamsdonk, 2010, 2011; Bianchi and Myers, 2014; Nomura et al., 2016).

5.1 Induced gravity

Induced gravity is a form of emergent gravity[25] that pursues the idea that gravitational dynamics originate in or are *induced* by quantum corrections from the matter fields. The programme originates in the discovery by Sakharov (1967) that semi-classical gravity in the form of the Einstein-Hilbert action necessarily occurs as part of the effective action in any quantum field theory on a Lorentzian manifold at first loop order, i.e. as first order quantum corrections to the classical equations of motion. In a few more words, the argument assumes a dynamical Lorentzian manifold with a generic QFT defined on it; nothing is assumed about the particulars of the field content of the QFT. The dynamics of the manifold is not specified other than that it couples to the matter content of the QFT. Based on this, one finds that the first order quantum corrections to the effective action – contributions from one-loop Feynman diagrams – necessarily include terms of the same form as the Einstein-Hilbert action from which the EFEs can be derived as equations of motion.[26] Thus, even if we do not have any gravitational dynamics in the form of EFEs or similar at tree level, i.e. at zero loops, it will occur at one-loop order in any QFT on a dynamical Lorentzian manifold. In a limit of the theory with one-loop dominance,[27] the effective action has the EFEs as equations of motion and the theory

---

[25]For a survey of emergent gravity, see Linnemann and Visser (2018).

[26]In the reconstruction by Visser (2002), the metric, g_{ab}, is expanded as the sum of a background, g_{ab}^0, and a perturbation, $h_{ab}(x)$ (nothing is assumed about the size of the perturbation). For the purposes of a Feynman diagram picture, the Fourier transform of the metric perturbation, $h_{ab}(k)$, takes the form of external gravitons. Sakharov's result is then that one-loop diagrams with an arbitrary number of external gravitons give rise to terms in the effective action of the form:

$$\int d^4x \sqrt{-g}\,[c_1 + c_2 R(g)]$$

where $R(g)$ is the Ricci tensor, and c_1, c_2 are dimensionless constants that depend on the particular particle species. Compare this to the Einstein-Hilbert action:

$$\int d^4x \sqrt{-g}\left[-\Lambda - \frac{R(g)}{16\pi G_N}\right]$$

where Λ is the cosmological constant. See Visser (2002, 5) for further details on the connection between these two actions.

[27]See Visser (2002) for a survey of different routes to such a limit.

therefore recovers general relativity according to the condition of section 2. Visser (2002) summarizes the implications in the following manner:

> If we consider any candidate theory for quantum gravity (brane models, quantum geometry, lattice quantum gravity) then Sakharov's scenario tells us that trying to derive the inverse square law from first principles is not the difficult step. Deriving the inverse square law, and all of Einstein gravity is in fact automatic once you have demonstrated the existence of Lorentzian manifolds (990).

The EFEs are generic equations of motion for QFTs on dynamical Lorentzian manifolds. Very little must be assumed about the details of the theory, for instance about the field content. Though it is confined to the QFT framework and dependent on the assumption of a Lorentzian manifold, Sakharov's scenario indicates, as signified by Visser, that the EFEs are generic in such theories. In other words, recovering general relativity by this route does not depend on the details of the underlying microstructure. This is the first indication that general relativity is multiply realizable. The point is not that induced gravity is the most promising approach to quantum gravity,[28] but rather that induced gravity signifies how generic the EFEs are.

5.2 Thermodynamic gravity

Jacobson's (1995) thermodynamic gravity[29] is similar to induced gravity in that it derives the EFEs from few general ingredients and without adding gravitational dynamics to the mix by hand. Thermodynamic gravity relies on two basic assumptions: a proportionality between entropy, S, and area and a Clausius-like relation[30] $\delta Q = T dS$ connecting heat, Q, and temperature, T, to entropy. In Jacobson's thermodynamic gravity, the heat flow, δQ, is qualified as the energy flow over the causal horizon as observed by a boosted observer, known as a Rindler horizon.[31] It is therefore given as an integral over the energy-momentum tensor contracted with the local boost Killing vector field generating the Rindler horizon, and an area element of the Rindler horizon. Since causal horizons hide information, they can be associated with an entropy.[32] The change in entropy, dS, associated with a piece of horizon is, by the assumed relation between entropy and area, proportional to the variation of the area of that piece of horizon. The change of area for a Rindler horizon can in turn be expressed as an integral of the Ricci tensor, the local boost Killing vector field generating the Rindler horizon and an area element

---

[28]Indeed, the Weinberg-Witten (1980) theorem suggests that induced gravity scenarios faces serious difficulties. See Sindoni (2012) for a discussion.

[29]There are many approaches going by the name of thermodynamic gravity all building to some extent on Jacobson's insight. See Padmanabhan (2010b) for a review.

[30]This is an equilibrium thermodynamic relation.

[31]The most well known example of a causal horizon is the event horizon of a black hole. The type of causal horizons considered here are those observed by a constantly accelerating observer as known from the Unruh (1976) effect.

[32]Again this is well known from black hole thermodynamics (Bekenstein, 1973).

of the horizon. Then, by inserting the integrals into the assumed equilibrium relation, $\delta Q = TdS$ and imposing energy-momentum conservation, the EFEs can be derived. Jacobson (1995) concludes: "Viewed in this way, the Einstein equation is an equation of state" (1260).

Evidently, a number of assumptions go into this derivation, and their truth is not directly the concern. Rather, what is of interest is that they are compatible with different types of microstructure, and thus sufficiently independent of the nature of the underlying microstructure. Heat, temperature, and entropy are all thermodynamic quantities related to the mean collective behaviour of underlying degrees of freedom.[33] They are macroscopic quantities related to macroscopic phenomena, visible only at a sufficiently coarse-grained level. In thermodynamic gravity, this insensitivity to the nature of the microstructure is inherited by the EFEs: "gravity is viewed as robust with respect to the underlying microstructure, in the sense that it only depends on coarse-grained notions such as entropy and temperature" (Linnemann and Visser, 2018, 4). Indeed, the central characteristic of equilibrium thermodynamics is that it is sufficiently insensitive to the individual microscopic degrees of freedom; for instance the momentum of individual particles: "thermodynamics was developed and used effectively decades before we understood the molecular structure of matter or its statistical mechanics" (Padmanabhan, 2010b, 31). This is signified by temperature – one of the central elements is Jacobson's derivation of the EFEs – being often portrayed as an exemplar of a multiply realizable function (e.g. Kim (1998); Bickle (2016)). Thermodynamics is multiply realizable.

An equilibrium thermodynamic description is only available if the macroscopic behaviour of a system is sufficiently robust with respect to the individual changes among the microscopic degrees of freedom; a thermodynamic equilibrium is achieved when "all reasonable macroscopic observables have steady values" (Pitowsky, 2006, 432).[34] Jacobson is well aware of this restriction to thermodynamic gravity: "It is born in the thermodynamic limit as a relation between thermodynamic variables, and its validity is seen to depend on the existence of local equilibrium conditions" (Jacobson, 1995). However, from the present perspective, this signifies that this approach to obtain the EFEs is genuinely thermodynamic. So long as the microscopic degrees of freedom have macroscopic equilibrium states[35] that admit a conception of heat in the form of energy transfer over causal horizons and entropy as causally hidden information that is related to the horizon area, then the EFEs can be obtained as thermodynamic equations of state and general relativity can as a consequence be recovered as a thermodynamic phenomenon. The existence of causal horizons in the form of Rindler horizons is simply guaranteed by the assumption in the approach of a Lorentzian manifold and the possible existence of boosted observers. The accelerated observer perceives a causal horizon and it is there-

---

[33]The rigorous foundation of the relation between the macroscopic thermodynamic level of description and the microstructure on which it supervenes remains elusive (Sklar, 1993; Ridderbos, 2002). Here it suffices to observe that thermodynamics somehow does work to make reliable predictions.

[34]This and many other aspects in the foundations of equilibrium thermodynamics are disputed. These will not be discussed further here. For a review, see Frigg (2008).

[35]See Sindoni (2013) for further details on how such an equilibration might obtain.

fore mandatory to integrate out the degrees of freedom causally hidden by the horizon which associates an entropy to the horizon. When matter crosses the horizon, then the entropy is changed according to the amount of energy transferred and the temperature of the reservoir which equates the temperature of the horizon: $\delta Q = TdS$.[36] The proportionality between entropy and horizon area is perhaps the assumption of Jacobson's thermodynamic gravity that looks the most restrictive and consequently the one that might hinder the EFEs from inheriting the versatility of thermodynamics. However, a proportionality between entropy and horizon area is known from black hole thermodynamics (Bekenstein, 1973). And the close relation between black hole and horizon thermodynamics suggests that such a proportionality should obtain for all causal horizons (Padmanabhan, 2005, section 7).[37] Black hole thermodynamics, in turn, is argued by Wallace (2018) to be genuinely thermodynamic,[38] and in so far as the proportionality between entropy and horizon area inherits this characteristic, it should also be sufficiently generic for thermodynamic gravity to inherit the multiple realizability of its thermodynamic origin:

> Just as the bulk properties of solids can be described without reference to the underlying atomic structure, much of classical and semi classical gravity (including the entropy of black holes) will be independent of the underlying description of the microscopic degrees of freedom (Padmanabhan, 2005, 114).

No special assumption has to be made about the nature of the microstructure that realizes thermodynamic gravity; it relies solely on very generic features. Thermodynamic gravity is as generic as other forms of thermodynamics, and presuming that many types of quantum gravity microstructure can realize thermodynamics, this shows a way in which general relativity can be multiply realizable.

In summary, this possible thermodynamic origin of general relativity corroborates the view that general relativity is multiply realizable in quantum gravity. The derivation shows how the EFEs can inherit the multiple realizability of temperature and other thermodynamic properties. Thermodynamics can be realized on systems with very different types of microstructure; presumably including those posited by theories of quantum

---

[36]See Padmanabhan (2010b, section 4) for further details. In essence, Jacobson's thermodynamic gravity derives from the thermodynamic nature of causal horizons. While horizon thermodynamics can be derived from the assumption that the EFEs are equations of state (e.g. Hansen et al. (2017)), horizon thermodynamics can also be defended on independent grounds and then shown to recover general relativity (Padmanabhan, 2005).

[37]This argument by analogy to the proportionality between entropy and horizon area is at risk of being circular since the derivation of black hole thermodynamics assumes gravity. We shall here ignore this concern for two reasons: First, since there are routes to thermodynamic gravity that do not assume a proportionality between entropy and horizon area (e.g. Padmanabhan (2010a)) and second, that the proportionality can be defended on independent grounds if the entropy is identified as entanglement entropy (more on this below). Observe that a similar concern arises if the proportionality is defended with reference to the holographic principle as formulated by 't Hooft (1994), Susskind (1995), and Bousso (2002) since this principle also ultimately relies on black hole thermodynamics.

[38]This view is disputed by Dougherty and Callender (2016).

gravity. Thus, since general relativity, by the argument above, can be realized as a thermodynamic effect, and thermodynamics is likely to be multiply realizable in quantum gravity, then so is general relativity. The recovery of general relativity in a theory of quantum gravity does not entail that the theory posits the right microstructure, since the EFEs are likely to be recoverable in many theories of quantum gravity that feature a Lorentzian manifold and a microstructure that admits thermodynamic equilibration.

5.3 Entanglement gravity

Already in his thermodynamic gravity, Jacobson alludes to the possibility that the entropy associated with the Rindler horizon might originate in entanglement.[39] Entanglement entropy measures how entangled the degrees of freedom within some subsystem are with the degrees of freedom outside it.[40] In the following, these subsystems will be associated with subregions of the manifold on which the QFT is defined. The first approach to entanglement gravity to be considered below assumes the AdS/CFT correspondence, whereas the second approach considered – also due to Jacobson – trades the assumption of AdS/CFT with an equilibrium assumption and some lessons from many-body systems.

The AdS/CFT correspondence is a conjectured duality between theories with gravity in asymptotically Anti-de Sitter spacetime (AdS)[41] and conformal field theories (CFT) defined on non-dynamical Lorentzian manifold (Maldacena, 1999), i.e. the matter fields of the CFT do not couple to the metric of the manifold and the metric does as a consequence not change.[42] Being a duality, the two theories are conjectured to be empirically equivalent, but different ways to represent the same physics.[43] In CFTs, a Clausius-like relation can be derived as a purely formal result: For any ball shaped region of a CFT, the change in entanglement entropy of a region is proportional to its change in hyper-

---

[39]It is disputed whether entanglement entropy is genuine thermodynamic entropy and the approaches to the EFEs based on entanglement entropy might as a consequence not be examples of *thermodynamic* gravity. There are indications that entanglement entropy is indeed connected to thermodynamic entropy (Kaufman et al., 2016) and regardless, the entanglement approaches to EFEs considered below at most assume an entanglement thermodynamics – genuine or not – analogous to ordinary thermodynamics (see Alishahiha et al. (2013) for more details on entanglement thermodynamics).

[40]Formally, the entanglement entropy, S_B of a quantum subsystem living in the manifold subregion B is defined as $S_B = -\rho_B \log(\rho_B)$. Here ρ_B is the reduced density matrix for the quantum subsystem.

[41]Anti-de Sitter spacetimes are vacuum solutions to the EFEs with a negative cosmological constant, i.e. solutions whose curvature is the other way around as compared to the current best cosmological models. It appears, therefore, that the AdS/CFT correspondence does not obtain in the actual world.

[42]The correspondence has its origin in string theory and conjectures more precisely that certain closed string theories in asymptotically AdS are dual to certain CFTs defined on a fixed background identical to the asymptotic boundary of the dual AdS spacetime. See Butterfield et al. (2016) for a philosophical introduction to the AdS/CFT correspondence.

[43]See for instance de Haro (2017) for more details about the notion of duality in the context of the AdS/CFT correspondence. All metaphysical questions, for instance whether the AdS side and the CFT side represent metaphysically distinct possible worlds or not will be set aside since both worlds—by definition of the notion of duality—would be empirically indistinguishable (see Read (2016) and Le Bihan and Read (2018) for an overview of these metaphysical questions).

bolic energy (Casini et al., 2011).[44] Since this is a formal result of CFTs, the relation holds always and not only for equilibrium states. In the limit where the AdS side can be approximated by semi-classical gravity, also a relation between entropy and areas can be shown to obtain. More precisely, this so-called Ryu-Takayanagi formula states that areas on the AdS side are proportional to the entanglement entropy of subsystems on the CFT side (Ryu and Takayanagi, 2006).[45] Using the Ryu-Takayanagi formula (and nothing else of the AdS/CFT correspondence), one can translate the CFT relation between entanglement entropy and hyperbolic energy to the AdS side where it proves to be equivalent to the linearized vacuum EFEs (Lashkari et al., 2014; Faulkner et al., 2014). Taking into account corrections to the Ryu-Takayanagi formula due to entanglement on the AdS side (Lewkowycz and Maldacena, 2013), the relation between entanglement entropy and hyperbolic energy instead translates into linearized EFEs with matter content and has indications of a generalization to the full EFEs (Swingle and Van Raamsdonk, 2014).

While still in need of further development, the argument suggests a way to conceive of the EFEs as entanglement dynamics and this sounds promising for the present purposes. Any quantum degrees of freedom can be entangled. Entanglement can, in other words, be realized in any quantum system with two or more degrees of freedom. Arguably, entanglement will, as a consequence, feature in any theory of *quantum* gravity regardless the details of its proposed microstructure; it just has to be quantum. Obviously, a number of assumptions go into the relation between the EFEs and entanglement, but if these can be shown to be sufficiently generic, then general relativity – by its assumed relation to the EFEs – might inherit the generic nature of entanglement in theories of quantum gravity. However, care must be taken when it comes to dualities. The result so far is that if the Ryu-Takayanagi formula with bulk entanglement corrections holds, then the EFEs (on the AdS side) can instead be encoded as a constraint on the relation between hyperbolic energy and entanglement entropy (on the CFT side). Several authors have argued that the AdS/CFT correspondence, as an exact duality, cannot testify to the *emergence* of gravity (e.g. Teh (2013), Rickles (2013), Dieks et al. (2015), and Butterfield et al. (2016)). The AdS side and CFT side are simply different ways to represent the same physics: "the presence of a duality precludes the phenomenon of emergence" (de Haro, 2017, 110). Just as we may conceive of the EFEs as entanglement dynamics in such circumstances, so it is possible to conceive of entanglement dynamics in terms of the EFEs. The situation, therefore, is different from induced and thermodynamic

---

[44]Formally we have $dS_B = \delta E_B^{hyp}$ if

$$\delta E_B^{hyp} = \pi \int_B d^{d-1}x \frac{R^2 - (\vec{x} - \vec{x_0})^2}{R} \delta \left\langle T^{tt}(x) \right\rangle$$

where R is the radius of the ball shaped region and $\delta \left\langle T^{tt}(x) \right\rangle$ is the variation of the energy density.

[45]In mathematical terms, the Ryu-Takayanagi formula takes the form: $S_B = \text{Area}(\tilde{B})/4G$ where S_B is the entanglement entropy of a subregion of the manifold, B, on the CFT side and $\tilde{B}$ is the area on the AdS side related to the region B on the CFT side. For further details about the Ruy-Takayanagi formula including the evidence for it within the AdS/CFT correspondence, see Rangamani and Takayanagi (2017).

gravity. These showed how the EFEs can be derived as an effective description holding in a particular limit – thermodynamic or first loop order – even if gravity is absent at the level of description of quantum gravity and thus only manifests itself as a large scale phenomenon. In these accounts, the EFEs are in this sense emergent. Such emergence is precluded for exact dualities and thus when the Ryu-Takayanagi formula relates areas of surfaces in one theory (the AdS side) to entanglement entropies in another theory (the CFT side). The Ryu-Takayanagi formula simply takes the form of a relation between representations like the entry of a dictionary. However, the concern here is not the emergence of gravity, but rather indications that general relativity is multiply realizable. If the EFEs are equivalent to some constraint on entanglement entropy and this constraint is multiply realizable, then this characteristic must, due to the duality, be shared by the EFEs and therefore general relativity: What is multiply realizable on one side cannot be a physical kind on the other.

As already argued, entanglement is to be expected in any theory of *quantum* gravity. However, the additional assumptions that go into entanglement gravity – the Ryu-Takayanagi formula and the relation between entanglement entropy and hyperbolic energy – are not as innocent. Assuming the full battery of the AdS/CFT correspondence, the relation between entanglement entropy and hyperbolic energy is a formal fact – it holds for all CFTs – and the Ryu-Takayanagi formula is proven to hold in the limit where the AdS side can be approximated by semi-classical gravity (Casini et al., 2011; Lewkowycz and Maldacena, 2013). The problem with this line of argument, however, is the AdS/CFT correspondence itself. The correspondence was originally conceived within string theory and is therefore deeply rooted in this particular approach to quantum gravity. In order for the EFEs to potentially inherit the versatility of entanglement, their relation cannot rely on idiosyncratic aspects of string theory. Thus, one must depart from string theory, and therefore the framework of the AdS/CFT correspondence, and seek more generic reasons for the Ryu-Takayanagi formula[46] and the relation between entanglement entropy and hyperbolic energy to hold.

One such attempt to establish a relation between the EFEs and entanglement inspired by the discoveries of AdS/CFT[47] but without stipulating the AdS/CFT correspondence is proposed by Jacobson (2016).[48] Jacobson also considers a ball shaped region of a Lorentzian manifold, but now with a generic QFT living on it. Without the AdS/CFT correspondence, the Ruy-Takayanagi formula is reinterpreted as a leading order propor-

---

[46]There are indications, though conjectural, that the Ryu-Takayanagi formula generalizes to other prospective theories of quantum gravity: causal set theory (Sorkin and Yazdi, 2018), loop quantum gravity (Smolin, 2016; Han and Hung, 2017), and the more general framework of generic group field theories employing a relation to tensor networks with promises of even further generalization (Chirco et al., 2018). But the worries raised in section 4 might be raised anew, if too much emphasis is put on our current field of prospective quantum gravity theories.

[47]Jacobsen explicitly refers to Lashkari et al. (2014) and Faulkner et al. (2014) as central sources of inspiration (2016, 1).

[48]For the sake of brevity, we shall here limit the attention to Jacobson (2016), but there are several other approaches to the EFEs from entanglement that are inspired by AdS/CFT without stipulating it, e.g. Verlinde (2017) and Cao and Carroll (2018).

tionality between the area and entanglement entropy of the ball. Jacobson motivates this proportionality at leading order with the observation that the primary contribution to the entanglement entropy comes from short range entanglement over the boundary of the ball which accounts for the area scaling.[49] The relation between entanglement entropy and hyperbolic energy is motivated by the assumption that the QFT under consideration has a UV fixed point such that it is conformally invariant at sufficiently small distances and that the ball is small enough to be conformally invariant throughout. Finally, due to the details of the argument, Jacobson must assume that the ball is at an entanglement equilibrium such that the first order variation of the entanglement entropy vanishes.[50] However, given these elements, the full EFEs can be derived. Thus, general relativity can be recovered as the result of entanglement equilibrium under the assumption of a UV fixed point and leading order area scaling of the entanglement entropy.

Being freed from the AdS/CFT correspondence and therefore string theory, Jacobson's derivation suggests a way for general relativity to inherit the generic nature of entanglement. Jacobson's assumed UV fixed point and entailed conformal invariance can be motivated as a necessary condition for any fundamental theory. This follows since the lack of a UV fixed point – often as connected with non-renormalizability – is usually taken as evidence that the theory is merely an effective theory that fails to incorporate some underlying microstructure.[51] As such, Jacobson's assumed UV fixed point should exist for any theory of quantum gravity that aims to be a fundamental theory. Also leading order area scaling of entanglement entropy is a common result, and seems for instance to be generic in quantum many body systems (Eisert et al., 2010). In addition, such area scaling suggests that the Bekenstein (1981) bound on entropy is saturated to leading order, something that accords well with the view that "holographic principle is generally viewed as an important guiding principle in the search for a theory of quantum gravity" (Linnemann and Visser, 2018).[52] Finally, the result that a small ball – a small local system – should be at an entanglement equilibrium after some time is supported by quantum thermalization in generic many-body systems (Kaufman et al., 2016). Here, it is found that small local systems over time thermalize, i.e. enter an equilibrium with respect to entanglement entropy even when the full system remains in a pure state. Both of these arguments from quantum many-body physics rely on macroscopic behaviour that – in analogy to the case of thermodynamics – promises to

---

[49]For the purposes of Jacobson's argument, it is even sufficient that the UV part of the entanglement entropy is proportional with area, an assumption that looks even less contestable.

[50]Jacobson also uses the (contracted) Bianchi identity which assumes that the connection is metric compatible. Though this is a very common assumption in general relativity, reality might prove to be otherwise, but we will not pursue this further here. Thank you to one of the anonymous reviewers of *Synthese* for raising this issue.

[51]Crowther and Linnemann (2017) argue that a theory of quantum gravity does not have to be fundamental and that UV-completeness cannot be assumed for quantum gravity. We will disregard this subtlety here.

[52]Some care must be taken when it comes to the holographic principle since its derivation assumes parts of general relativity. Using it to argue that general relativity is multiply realizable might therefore be circular. This, however, it not the place to discuss this issue further. See (Bousso, 2002, section ii) for a discussion of the various routes to the holographic principle.

be sufficiently insensitive to the details of the underlying microstructure; their nature is unimportant for the mentioned effects.

In summary, Jacobson's derivation suggests how the EFEs can obtain as an effective description due to generic entanglement dynamics satisfying the entanglement equilibrium and leading order area scaling of entanglement. With these elements of the derivation and entanglement itself promising to be generic features in quantum gravity, this again entails that general relativity can be multiply realizable in quantum gravity. The recovery of general relativity in quantum gravity theories is unsurprising due to this relation to entanglement that in turn should feature in any theory of *quantum* gravity.

6   Conclusion

Does such a recovery of general relativity in a theory of quantum gravity ensure that the microstrcuture it posits finds a recognizable counterpart in the true theory of quantum gravity? Based on a functionalist perspective, this paper has defended a negative answer to the question: General relativity can be recovered from elements that are multiply realizable or at least generic in nature in quantum gravity, and one can therefore expect that general relativity can be recovered in theories of quantum gravity that describe other possible worlds than the actual one. A theory of quantum gravity can recover general relativity, while being completely wrong about the microstructure it posits.

The argument was founded on three research schemes – induced gravity, thermodynamic gravity, and entanglement gravity – where the EFEs are derived from generic elements that do not invoke any particulars of the type of microstructure realizing the EFEs. With the insensitivity to the microstructure, these accounts suggest how the EFEs can be multiply realizable. With reference to string theory, it was indicated what is meant, when a theory of quantum gravity is said to recover general relativity, and the three accounts each provide an indication of how general relativity can be multiply realizable in quantum gravity. Given the character of each case study, I conjecture that we can expect many prospective theories of quantum to recover general relativity even though the microstructure they posit are not that of the actual world.

This impedes to a degree the role of general relativity in non-empirical confirmation of quantum gravity. If general relativity, as argued here, is multiply realizable in quantum gravity, then we should not be too impressed by a theory of quantum gravity where general relativity can be recovered in some limit of the theory. The recovery of general relativity cannot serve as the ultimate arbiter that decides which theory of quantum gravity that is worthy of pursuit, even though it is of course not irrelevant either *qua* quantum *gravity*. Returning where we started, it is arguably essential that string theory recovers general general relativity for its prospects as a theory of quantum gravity, but it does not establish that string theory must have some truth to it.

References

Mohsen Alishahiha, Davood Allahbakhshi, and Ali Naseh. Entanglement thermodynamics. *Journal of High Energy Physics*, 2013(8):102, August 2013. ISSN 1029-8479. doi: 10.1007/JHEP08(2013)102. URL https://doi.org/10.1007/JHEP08(2013)102.

I. Antoniadis, E. Kiritsis, J. Rizos, and T.N. Tomaras. D-branes and the Standard Model. *Nuclear Physics B*, 660(1):81–115, June 2003. ISSN 0550-3213. doi: 10.1016/S0550-3213(03)00256-6. URL http://www.sciencedirect.com/science/article/pii/S0550321303002566.

Carlos Barceló, Stefano Liberati, and Matt Visser. Analogue Gravity. *Living Reviews in Relativity*, 14(1):3, May 2011. ISSN 1433-8351. doi: 10.12942/lrr-2011-3. URL https://doi.org/10.12942/lrr-2011-3.

Bekir Baytaş, Eugenio Bianchi, and Nelson Yokomizo. Gluing polyhedra with entanglement in loop quantum gravity. *Physical Review D*, 98(2):026001, 2018.

George Bealer. Self-Consciousness. *The Philosophical Review*, 106(1):69–117, 1997. ISSN 00318108, 15581470. doi: 10.2307/2998342. URL http://www.jstor.org/stable/2998342.

Katrin Becker, Melanie Becker, and John H Schwarz. *String theory and M-theory : a modern introduction*. Cambridge University Press, Cambridge, 2007. ISBN 0-521-86069-5.

Jacob D. Bekenstein. Black Holes and Entropy. *Physical Review D*, 7(8):2333–2346, 1973.

Jacob D. Bekenstein. Universal upper bound on the entropy-to-energy ratio for bounded systems. *Physical Review D*, 23(2):287–298, January 1981.

Eugenio Bianchi and Robert C. Myers. On the architecture of spacetime geometry. *Classical and Quantum Gravity*, 31(21):214002, 2014.

John Bickle. Multiple Realizability, 2016. URL https://plato.stanford.edu/archives/spr2017/entries/mind-identity/.

Raphael Bousso. The holographic principle. *Reviews of Modern Physics*, 74(3):825–874, 2002.

Harvey R. Brown. *Physical Relativity: Space-time structure from a dynamical perspective*. Oxford University Press, Oxford, 2005. ISBN 978-0-19-927583-0.

Jeremy Butterfield, Sebastian de Haro, and Daniel R. Mayerson. Conceptual Aspects of Gauge/Gravity Duality. *Foundations of Physics*, 46(11):1381–1425, 2016.

Frank Cabrera. String theory, non-empirical theory assessment, and the context of pursuit. *Synthese*, October 2018. ISSN 1573-0964. doi: 10.1007/s11229-018-01987-9. URL https://doi.org/10.1007/s11229-018-01987-9.

C.G. Callan, D. Friedan, E.J. Martinec, and M.J. Perry. Strings in background fields. *Nuclear Physics B*, 262(4):593–609, December 1985. ISSN 0550-3213. doi: 10.1016/0550-3213(85)90506-1. URL `http://www.sciencedirect.com/science/article/pii/0550321385905061`.

C.G. Callan, C. Lovelace, C.R. Nappi, and S.A. Yost. String loop corrections to beta functions. *Nuclear Physics B*, 288:525–550, 1987.

Kristian Camilleri and Sophie Ritson. The role of heuristic appraisal in conflicting assessments of string theory. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 51:44–56, August 2015. ISSN 1355-2198. doi: 10.1016/j.shpsb.2015.07.003. URL `http://www.sciencedirect.com/science/article/pii/S1355219815000490`.

ChunJun Cao and Sean M. Carroll. Bulk entanglement gravity without a boundary: Towards finding Einstein's equation in Hilbert space. *Physical Review D*, 97(8):086003, April 2018. URL `https://link.aps.org/doi/10.1103/PhysRevD.97.086003`.

Steven Carlip. Challenges for emergent gravity. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 46, Part B:200–208, 2014.

Horacio Casini, Marina Huerta, and Robert C. Myers. Towards a derivation of holographic entanglement entropy. *Journal of High Energy Physics*, 2011(5):36, 2011.

Goffredo Chirco, Hal M. Haggard, Aldo Riello, and Carlo Rovelli. Spacetime thermodynamics without hidden degrees of freedom. *Physical Review D*, 90(4):044044, August 2014.

Goffredo Chirco, Daniele Oriti, and Mingyi Zhang. Group field theory and tensor networks: towards a Ryu–Takayanagi formula in full quantum gravity. *Classical and Quantum Gravity*, 35(11):115011, 2018.

Karen Crowther and Niels Linnemann. Renormalizability, Fundamentality, and a Final Theory: The Role of UV-Completion in the Search for Quantum Gravity. *The British Journal for the Philosophy of Science*, pages axx052–axx052, September 2017. ISSN 0007-0882. doi: 10.1093/bjps/axx052. URL `http://dx.doi.org/10.1093/bjps/axx052`.

Karen Crowther, Niels S. Linnemann, and Christian Wüthrich. What we cannot learn from analogue experiments. *Synthese*, May 2019. ISSN 1573-0964. doi: 10.1007/s11229-019-02190-0. URL `https://doi.org/10.1007/s11229-019-02190-0`.

Radin Dardashti, Karim P. Y. Thébault, and Eric Winsberg. Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity. *The British Journal for the Philosophy of Science*, 68(1):55–89, May 2015. ISSN 0007-0882. doi: 10.1093/bjps/axv010. URL `https://doi.org/10.1093/bjps/axv010`.

Radin Dardashti, Richard Dawid, and Karim Thébault, editors. *Why Trust a Theory?: Epistemology of Fundamental Physics*. Cambridge University Press, Cambridge, 2019. ISBN 978-1-108-47095-7. doi: 10.1017/9781108671224. URL `https://www.cambridge.org/core/books/why-trust-a-theory/ABE64219FA50E2546F5C72B8168A10C0`.

Richard Dawid. Underdetermination and Theory Succession from the Perspective of String Theory*. *Philosophy of Science*, 73(3):298–322, 2006. ISSN 00318248, 1539767X. doi: 10.1086/515415. URL `http://www.jstor.org/stable/10.1086/515415`.

Richard Dawid. On the Conflicting Assessments of the Current Status of String Theory. *Philosophy of Science*, 76(5):984–996, 2009. ISSN 00318248, 1539767X. doi: 10.1086/605794. URL `http://www.jstor.org/stable/10.1086/605794`.

Richard Dawid. *String Theory and the Scientific Method.* Cambridge University Press, Cambridge, 2013a.

Richard Dawid. Theory Assessment and Final Theory Claim in String Theory. *Foundations of Physics*, 43(1):81–100, January 2013b. ISSN 1572-9516. doi: 10.1007/s10701-011-9592-x. URL `https://doi.org/10.1007/s10701-011-9592-x`.

Sebastian de Haro. Dualities and emergent gravity: Gauge/gravity duality. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 59:109–125, 2017.

Dennis Dieks, Jeroen van Dongen, and Sebastian de Haro. Emergence in holographic scenarios for gravity. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 52, Part B:203–216, 2015.

John Dougherty and Craig Callender. Black Hole Thermodynamics: More Than an Analogy? 2016. URL `http://philsci-archive.pitt.edu/13195/`.

Fay Dowker. Causal sets as discrete spacetime. *Contemporary Physics*, 47(1):1–9, January 2006.

J. Eisert, M. Cramer, and M. B. Plenio. Colloquium: Area laws for the entanglement entropy. *Reviews of Modern Physics*, 82(1):277–306, February 2010. doi: 10.1103/RevModPhys.82.277. URL `https://link.aps.org/doi/10.1103/RevModPhys.82.277`.

Thomas Faulkner, Monica Guica, Thomas Hartman, Robert C. Myers, and Mark Van Raamsdonk. Gravitation from entanglement in holographic CFTs. *Journal of High Energy Physics*, 2014(3):51, 2014.

Roman Frigg. A Field Guide to Recent Work on the Foundations of Statistical Mechanics. In Dean Rickles, editor, *The Ashgate Companion to Contemporary Philosophy of Physics*. Ashgate, 2008.

Brian Greene. *The elegant universe : superstrings, hidden dimensions, and the quest for the ultimate theory.* W.W. Norton, New York, 1999. ISBN 0-393-04688-5.

Muxin Han. Einstein equation from covariant loop quantum gravity in semiclassical continuum limit. *Physical Review D*, 96(2):024047, July 2017. doi: 10.1103/PhysRevD.96.024047. URL `https://link.aps.org/doi/10.1103/PhysRevD.96.024047`.

Muxin Han and Ling-Yan Hung. Loop quantum gravity, exact holographic mapping, and holographic entanglement entropy. *Physical Review D*, 95(2):024011, 2017.

Devin Hansen, David Kubizňák, and Robert B. Mann. Horizon thermodynamics from Einstein's equation of state. *Physics Letters B*, 771:277–280, August 2017. ISSN 0370-2693. doi: 10.1016/j.physletb.2017.04.076. URL `http://www.sciencedirect.com/science/article/pii/S0370269317303507`.

Nick Huggett. Target space is not equal to space. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 59:81–88, 2017.

Nick Huggett and Tiziana Vistarini. Deriving General Relativity from String Theory. *Philosophy of Science*, 82(5):1163–1174, 2015.

Nick Huggett and Christian Wüthrich. Emergent spacetime and empirical (in)coherence. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44 (3):276–285, 2013.

Steven Jackson, Lauren McGough, and Herman Verlinde. Conformal bootstrap, universality and gravitational scattering. *Nuclear Physics B*, 901:382–429, December 2015.

Ted Jacobson. Thermodynamics of Spacetime: The Einstein Equation of State. *Physical Review Letters*, 75(7):1260–1263, 1995.

Ted Jacobson. Entanglement Equilibrium and the Einstein Equation. *Physical Review Letters*, 116(20): 201101, 2016.

Adam M. Kaufman, M. Eric Tai, Alexander Lukin, Matthew Rispoli, Robert Schittko, Philipp M. Preiss, and Markus Greiner. Quantum thermalization through entanglement in an isolated many-body system. *Science*, 353(6301):794, August 2016. doi: 10.1126/science.aaf6725. URL `http://science.sciencemag.org/content/353/6301/794.abstract`.

Jaegwon Kim. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, volume 75. MIT Press, 1998.

Eleanor Knox. Effective spacetime geometry. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44(3):346–356, August 2013. ISSN 1355-2198. doi: 10.1016/j.shpsb.2013.04.002. URL `http://www.sciencedirect.com/science/article/pii/S1355219813000373`.

Eleanor Knox. Physical relativity from a functionalist perspective. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, October 2017. ISSN 1355-2198. doi: 10.1016/j.shpsb.2017.09.008. URL `http://www.sciencedirect.com/science/article/pii/S1355219817301247`.

Vincent Lam and Michael Esfeld. A dilemma for the emergence of spacetime in canonical quantum gravity. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44(3):286–293, 2013.

Vincent Lam and Christian Wüthrich. Spacetime is as spacetime does. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, June 2018. ISSN 1355-2198. doi: 10.1016/j.shpsb.2018.04.003. URL `http://www.sciencedirect.com/science/article/pii/S1355219817301703`.

Nima Lashkari, Michael B. McDermott, and Mark Van Raamsdonk. Gravitational dynamics from entanglement "thermodynamics". *Journal of High Energy Physics*, 2014(4):195, 2014.

Baptiste Le Bihan. Space Emergence in Contemporary Physics: Why We Do Not Need Fundamentality, Layers of Reality and Emergence. *Disputatio*, 10(49):71–95, 2018. doi: 10.2478/disp-2018-0004. URL `https://content.sciendo.com/view/journals/disp/10/49/article-p71.xml`.

Baptiste Le Bihan and James Read. Duality and ontology. *Philosophy Compass*, 13(12):e12555, December 2018. URL `https://doi.org/10.1111/phc3.12555`.

Aitor Lewkowycz and Juan Maldacena. Generalized gravitational entropy. *Journal of High Energy Physics*, 2013(8):90, 2013.

Niels S. Linnemann and Manus R. Visser. Hints towards the emergent nature of gravity. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 64:1–13, November 2018. ISSN 1355-2198. doi: 10.1016/j.shpsb.2018.04.001. URL `http://www.sciencedirect.com/science/article/pii/S1355219817302046`.

Juan Maldacena. The Large-N Limit of Superconformal Field Theories and Supergravity. *International Journal of Theoretical Physics*, 38(4):1113–1133, 1999.

Juan Maldacena and Leonard Susskind. Cool horizons for entangled black holes. *Fortschritte der Physik*, 61(9):781–811, 2013.

Thomas G. Mertens, Gustavo J. Turiaci, and Herman L. Verlinde. Solving the Schwarzian via the conformal bootstrap. *Journal of High Energy Physics*, 2017(8):136, August 2017.

Yasunori Nomura, Nico Salzetta, Fabio Sanches, and Sean J. Weinberg. Spacetime equals entanglement. *Physics Letters B*, 763:370–374, December 2016. ISSN 0370-2693. doi: 10.1016/j.physletb.2016.10.045. URL http://www.sciencedirect.com/science/article/pii/S0370269316306165.

Daniele Oriti. Disappearance and emergence of space and time in quantum gravity. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 46:186–199, May 2014.

T. Padmanabhan. Gravity and the thermodynamics of horizons. *Physics Reports*, 406(2):49–125, January 2005. ISSN 0370-1573. doi: 10.1016/j.physrep.2004.10.003. URL http://www.sciencedirect.com/science/article/pii/S0370157304004582.

T. Padmanabhan. EQUIPARTITION OF ENERGY IN THE HORIZON DEGREES OF FREEDOM AND THE EMERGENCE OF GRAVITY. *Modern Physics Letters A*, 25(14):1129–1136, May 2010a. ISSN 0217-7323. doi: 10.1142/S021773231003313X. URL https://doi.org/10.1142/S021773231003313X.

T. Padmanabhan. Thermodynamical aspects of gravity: new insights. *Reports on Progress in Physics*, 73(4):046901, 2010b.

Itamar Pitowsky. On the definition of equilibrium. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 37(3):431–438, September 2006. ISSN 1355-2198. doi: 10.1016/j.shpsb.2006.03.001. URL http://www.sciencedirect.com/science/article/pii/S1355219806000281.

Joseph Polchinski. *String theory : Vol. 1 : An introduction to the bosonic string*, volume Vol. 1. Cambridge University Press, Cambridge, 1998. ISBN 0-521-63303-6.

Mukund Rangamani and Tadashi Takayanagi. *Holographic entanglement entropy*. Springer, Cham, Switzerland, 2017.

James Read. The Interpretation of String-Theoretic Dualities. *Foundations of Physics*, 46(2):209–235, February 2016. ISSN 1572-9516.

Dean Rickles. AdS/CFT duality and the emergence of spacetime. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44(3):312–320, 2013.

Katinka Ridderbos. The coarse-graining approach to statistical mechanics: how blissful is our ignorance? *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 33(1):65–77, March 2002. ISSN 1355-2198. doi: 10.1016/S1355-2198(01)00037-5. URL http://www.sciencedirect.com/science/article/pii/S1355219801000375.

Carlo Rovelli. A Critical Look at Strings. *Foundations of Physics*, 43(1):8–20, January 2013. ISSN 1572-9516. doi: 10.1007/s10701-011-9599-3. URL https://doi.org/10.1007/s10701-011-9599-3.

Shinsei Ryu and Tadashi Takayanagi. Holographic Derivation of Entanglement Entropy from the anti-de Sitter Space/Conformal Field Theory Correspondence. *Phys. Rev. Lett.*, 96(18):181602, 2006.

Andrei D Sakharov. Vacuum quantum fluctuations in curved space and the theory of gravitation. *Dokl. Akad. Nauk SSSR*, 177(1):70–71, 1967.

Lorenzo Sindoni. Emergent Models for Gravity: an Overview of Microscopic Models. *SIGMA*, 8:027, 2012.

Lorenzo Sindoni. Horizon thermodynamics in pregeometry. *Journal of Physics: Conference Series*, 410 (1):012140, 2013. ISSN 1742-6596. URL `http://stacks.iop.org/1742-6596/410/i=1/a=012140`.

Lawrence Sklar. *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge University Press, Cambridge, 1993. ISBN 978-0-521-55881-5. doi: 10. 1017/CBO9780511624933. URL `https://www.cambridge.org/core/books/physics-and-chance/12F025B78A8AB1F840D63A51C409E219`.

J. J. C. Smart. The Mind/Brain Identity Theory, 2017. URL `https://plato.stanford.edu/archives/spr2017/entries/mind-identity/`.

Lee Smolin. *The trouble with physics : the rise of string theory, the fall of a science, and what comes next*. Houghton Mifflin, Boston, 2006.

Lee Smolin. Holographic relations in loop quantum gravity. 2016. URL `https://arxiv.org/abs/1608.02932`.

Rafael D Sorkin and Yasaman K Yazdi. Entanglement entropy in causal set theory. *Classical and Quantum Gravity*, 35(7):074004, 2018.

Andrew Strominger and Cumrun Vafa. Microscopic origin of the Bekenstein-Hawking entropy. *Physics Letters B*, 379(1):99–104, June 1996. ISSN 0370-2693. doi: 10.1016/0370-2693(96)00345-0. URL `http://www.sciencedirect.com/science/article/pii/0370269396003450`.

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

Brian Swingle and Mark Van Raamsdonk. Universality of Gravity from Entanglement. 2014. URL `https://arxiv.org/abs/arXiv:1405.2933v1`.

Gerard 't Hooft. Dimensional reduction in quantum gravity. In A. Ali, J. Ellis, S. Randjbar-Daemi, S. Weinberg, and Yu Lu, editors, *Highlights of Particle and Condensed Matter Physics (Salamfest): Proceedings of the Conference, ITCP, Trieste, Italy, 1993*, number 4 in World Scientific series in 20th Cenutry Physics, pages 284–296. World Scientific Publishing Co Pte Ltd, 1994.

Nicholas J. Teh. Holography and emergence. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44(3):300–311, 2013.

W. G. Unruh. Notes on black-hole evaporation. *Physical Review D*, 14(4):870–892, 1976.

Mark Van Raamsdonk. Building up spacetime with quantum entanglement. *General Relativity and Gravitation*, 42(10):2323–2329, 2010.

Mark Van Raamsdonk. A patchwork description of dual spacetimes in AdS/CFT. *Classical and Quantum Gravity*, 28(6):065002, 2011.

Erik Verlinde. On the origin of gravity and the laws of Newton. *Journal of High Energy Physics*, 2011 (4):29, 2011.

Erik Verlinde. Emergent Gravity and the Dark Universe. *SciPost Physics*, 3(2):016, 2017.

Matt Visser. SAKHAROV'S INDUCED GRAVITY: A MODERN PERSPECTIVE. *Modern Physics Letters A*, 17(15n17):977–991, June 2002. ISSN 0217-7323. doi: 10.1142/S0217732302006886. URL `https://doi.org/10.1142/S0217732302006886`.

David Wallace. The case for black hole thermodynamics part I: Phenomenological thermodynamics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, June 2018. ISSN 1355-2198. doi: 10.1016/j.shpsb.2018.05.002. URL `http://www.sciencedirect.com/science/article/pii/S1355219817301661`.

Steven Weinberg and Edward Witten. Limits on massless particles. *Physics Letters B*, 96(1):59–62, October 1980. ISSN 0370-2693. doi: 10.1016/0370-2693(80)90212-9. URL `http://www.sciencedirect.com/science/article/pii/0370269380902129`.

Silke Weinfurtner, Edmund W. Tedford, Matthew C. J. Penrice, William G. Unruh, and Gregory A. Lawrence. Classical Aspects of Hawking Radiation Verified in Analogue Gravity Experiment. In Daniele Faccio, Francesco Belgiorno, Sergio Cacciatori, Vittorio Gorini, Stefano Liberati, and Ugo Moschella, editors, *Analogue Gravity Phenomenology: Analogue Spacetimes and Horizons, from Theory to Experiment*, pages 167–180. Springer International Publishing, Cham, 2013.

Michael Weisberg. Robustness Analysis. *Philosophy of Science*, 73(5):730–742, 2006. ISSN 00318248, 1539767X. doi: 10.1086/518628. URL `http://www.jstor.org/stable/10.1086/518628`.

William Wimsatt. Robustness, Reliability and Overdetermination. In M. Brewer and B. Collins, editors, *Scientific Inquiry and the Social Sciences*, pages 124–163. Jossey-Bass, San Francisco, 1981.

Christian Wüthrich. Raiders of the Lost Spacetime. In Dennis Lehmkuhl, Gregor Schiemann, and Erhard Scholz, editors, *Towards a Theory of Spacetime Theories*, pages 297–335. Springer New York, New York, NY, 2017.

Barton Zwiebach. *A first course in string theory*. Cambridge University Press, Cambridge, 2nd edition, 2009.