Jointly-supervised PhD Thesis

University of Urbino *Carlo Bo*

Department of Pure and Applied Sciences

University of Geneva

Department of Philosophy

PhD Program in Research Methods in Science and Technology

XXXV Cycle

Breaking the Fabric: on AdS/CFT, Humeanism, and the Fate of Spacetime Inside Black Holes

Scientific Disciplinary Sector: M-FIL/02

PhD Candidate

Enrico Cinti

Supervisors

Prof. Vincenzo Fano (University of Urbino)

Prof. Christian Wüthrich (University of Geneva)

Academic Year: 2021/2022

Yet from those flames

No light, but rather darkness visible

(Milton, 1674, Book 1, lines 62-63)

Acknowledgements

This thesis would not have been possible without the help and guidance of my two supervisors, Vincenzo Fano and Christian Wüthrich. To them goes my deepest gratitude for the countless hours spent helping me become a better philosopher, and hopefully leading to a satisfactory work.

I owe an equally big debt of gratitude to my family, and especially my parents Patrizia and Stefano, and my brother Giorgio, for their continued support and encouragement. I would not be here if it was not for them.

Finally, I wish to thank the Synergia research group and the Geneva Symmetry Group, my two intellectual homes during this PhD, and the people involved with these groups, for many hours of invaluable intellectual discussion. Working together with the wonderful people of these groups has been the real highlight of my PhD experience.

Abstract

This thesis discusses the philosophy behind models of quantum black holes in AdS/CFT. These models display a variety of foundationally and metaphysically interesting features, ranging from modifications to the semiclassical structure of spacetime, to wide-ranging impacts on the metaphysics of laws of nature and the nature of spacetime emergence in Quantum Gravity.

In particular, in this thesis I start by analyzing the resolution of the firewall paradox within these models, and argue that it depends on the introduction of certain non-local connections within the spacetime structure of General Relativity. I then discuss the relation between these constructions and the topic of causality conditions on evaporating black holes in General Relativity, and suggest that various theorems apparently contradicting these AdS/CFT models of black holes do not apply due to, also in this case, subtle modifications of the standard spacetime structure of General Relativity.

On the metaphysics side, I use first of all these AdS/CFT models of quantum black hole to illustrate how Humeanism can be articulated in Quantum Gravity, suggesting in particular that spacetime relations should be substituted

with amplitude relations to do so. I also argue that in AdS/CFT, black holes make standard approaches to spacetime emergence untenable. Indeed, my contention is that spacetime emergence does not make sense in these cases, and that spacetime itself does not exist. Rather, all we have at our disposal is the operational data associated to spacetime, and I argue that this is sufficient to avoid issues having to do with the disappearance of spacetime in Quantum Gravity. Finally, I argue that, given this understanding of how classical spacetime is represented in AdS/CFT, we should prefer a Humean metaphysics in regard to laws of nature, at least insofar as treating General Relativity as a scientific theory involving laws of nature is a desideratum.

Contents

Chapter 1

Introduction

In recent years, there have been significant advances in AdS/CFT, a duality relating quantum gravity (QG) in Anti de Sitter (AdS) spacetime to a conformal field theory (CFT) in one dimension less (Maldacena, 1999; Witten, 1998) (see De Haro et al. (2016) for a philosophical introduction to AdS/CFT).[1] These advances mostly revolve around the quantum description of black holes within AdS/CFT, and have lead to a variety of surprising and previously unexpected results, among which:

- The derivation of the Page Curve, one of the main signals that the underlying QG theory is unitary (Penington, 2020; Almheiri et al., 2019). This result hints to an unitarity-based resolution of the black hole information loss paradox, whose original formulation is due to

[1]Technical jargon used throughout this introduction will be explained in detail throughout the thesis. For the purposes of this introduction, the rough intuitive glosses presented here will be sufficient.

Hawking (1976).

- The resolution of a variety of paradoxes surrounding unitary black hole evaporation (Maldacena and Susskind, 2013; Papadodimas and Raju, 2016; Penington, 2020), in particular paradoxes related and stemming from the work of Almheiri et al. (2013) around the firewall paradox. Indeed, the firewall paradox can be seen as the latest, most modern incarnation of Hawking's information loss paradox, and as such as carrying the essence of what is paradoxical about unitary black hole evaporation.

- The realization that the duality map relating (the yet unknown theory of) QG in AdS to the CFT, i.e., the map instantiating the duality between AdS and CFT in AdS/CFT, cannot be an isometry (Akers et al., 2022). Not being an isometry, in this context, means that the dimension of the fundamental QG Hilbert space is much smaller than the dimension of the Hilbert space of semiclassical gravity. Hence, semiclassical gravity overcounts the degrees of freedom of QG. This fact is at the heart of many of the black hole paradoxes.

- The semiclassical gravitational path integral can be used to derive the majority of these results, thus removing any specific AdS/CFT input, at least in principle (Penington et al., 2019; Almheiri et al., 2020). Nonetheless, to do this certain non-standard geometries have to be included in the path integral, leading to a modification of the standard result that there is no Page curve in semiclassical gravity. These geometries, known as replica wormholes, seem to be specific to gravity, in the sense that only a gravitational theory seems to give rise to them.

- Concrete experiments have been proposed to test the basic theoretical framework behind these models (Brown et al., 2019; Nezami et al., 2021), relying mostly on their possible implications for the structure of quantum information in the dual CFT. Hence, these are not simply theoretical speculations, but there is active experimental research into testing their basic features. Indeed, some preliminary results have been obtained (Jafferis et al., 2022) which suggest that these features are indeed present in concrete testable models.

- Lastly, many of these results seem to suggest that a new picture of geometry emerges from QG, where quantum effects are geometrized and change non trivially the standard geometric picture that general relativity (GR) originally afforded us. These new geometries have been suggested to be in the spirit of the ER=EPR proposal of Maldacena and Susskind (2013), which conjectures that there is a correspondence in QG between entanglement and space-time wormholes.

Given this wealth of fascinating and exciting results, it seems high time for philosophers to try to analyze their ultimate import. The goal of this thesis is to do exactly this, and to explore the philosophical significance and import of these recent results coming from the quantum treatment of black holes within AdS/CFT. In particular, the thesis addresses this issue from two different points of view. First, in the first two chapters, the problem of making sense of these results is approached from a more foundational perspective, with the goal of understanding what is the basic conceptual picture behind these AdS/CFT models of black holes. To accomplish this, I focus on two topics: the first is identifying the basic conceptual strategy used in these constructions to resolve paradoxes such as the firewall paradox of Almheiri et al. (2013); I argue that this is accomplished by a clever modification of the basic principle governing locality in non-QG physics, and hence of GR's spacetime structure. The second topic is the relevance of various theorems in GR showing that black hole evaporation cannot be causal for the purported derivation of the Page curve that these models afford. Here, too, my argument is that subtle departures from the spacetime structure of GR allow these AdS/CFT models of black holes to avoid theorems showing that black hole evaporation cannot be causal.

The conclusion of this first part then is that the changes to spacetime structure that QG engenders, while small (at least in the semiclassical regime), are nonetheless crucial to understand the functioning of black holes in AdS/CFT, and the results described above, which would apparently contradict a variety of theorems and paradoxes. The way these theorems and paradoxes are circumvented, as I argue in the thesis, is that while they assume that a standard GR spacetime is a good background to reason about black holes in QG, this assumption is violated in these specific constructions in AdS/CFT.

The last three chapters move from the foundational to a more metaphysics oriented approach, where the goal is to understand the implications of these models, and of their possible correctness, for a variety of questions in metaphysics. In particular, I focus on issues having to do with Humeanism and with space-time emergence, which in this context turn out to be tightly connected.

In particular, I argue in chapter 4 that these models can be used to give a formulation of Humeanism that avoids reliance in a background spacetime, as in Lewis (1994)'s classical formulation, and hence is compatible, contra the arguments of Lam and Wüthrich (2021a), with QG's apparent pronouncement that spacetime fundamentally does not exist. Moreover, I argue in chapter 5 that these models provide an extremely radical example of how spacetime might be related to the fundamental QG degrees of freedom. In particular, I argue that in these AdS/CFT models only the operational content of spacetime makes sense, and nothing else. Hence, space-time is neither emergent, nor exists, in these AdS/CFT constructions. Rather, only the operational data associated with a spacetime theory like GR makes sense, and nothing else. Finally, in chapter 6 I argue that this specific forum of the relation (or lack thereof) between spacetime and QG degrees of freedom can only be made sense within a Humean framework, at least under the assumption that one wants to count GR as involving laws of nature. A more modally committed framework cannot make sense of the extremely weak form of spacetime structure that these AdS/CFT models of black holes appear to be suggesting.

The basic methodology employed throughout the thesis is that of naturalized metaphysics (Ladyman et al., 2007; French, 2014), understood here as a style of philosophical inquiry that privileges attention to concrete scientific theories and their content instead of a priori reasoning. This approach in particular is applied by Ladyman et al. (2007) to the realm of metaphysical inquiry, where they suggest that metaphysical theses should always be grounded in the relevant scientific theories and should depend, whenever possible, on empirical theses liable to experimental testing. As such, metaphysics is transformed from a purely a priori enterprise into one that is grounded in scientific inquiry, and that extends as much as possible the methods of science to philosophy, including empirical testing.

By adopting a naturalized metaphysics perspective, I take as principal aims of my work the elucidation of the content of scientific theories, both along foundational lines as in the first part of the thesis, and along more metaphysical lines as in the second part of the thesis. This means that my main preoccupation in this thesis will be with the actual import and intimation of the scientific theories that I am analyzing, rather than with a priori beliefs about how reality should be. Indeed, even when dealing with more classically metaphysical topics such as Humeanism, I will strive to ground their discussion in the actual content of the relevant scientific theories. As such, I will not discuss for example Humeanism per se, but rather use Humeanism as a tool to better understand the consequences of certain scientific hypotheses about the world, as encoded in the AdS/CFT models I will discuss.

An important point, that I will stress at various junctures throughout the thesis, is that the reliance on concrete constructions in AdS/CFT for my analysis, and thus its model dependence, should not be seen, from the perspective of naturalized metaphysics, as a bug, but rather as a feature. Indeed, this model dependence entails that especially the metaphysics discussion in the thesis non trivially depends on the truth of certain empirically testable hypothesis, i.e., those encoded in the AdS/CFT models I discuss, one of Ladyman et al. (2007)'s condition for something to count as naturalized metaphysics. Depending on the results of the proposed experiments that I mentioned above, we might have reasonably soon direct empirical evidence in favor or against these models, and hence against the metaphysical hypothesis grounded on them that I put forward in this thesis. Hence, these proposals can be counted as properly naturalistic, science based metaphysics proposals, in the sense articulated in Ladyman et al. (2007).

Before moving to a slightly more detailed description of each chapter, let me address a possible objection one might have to the overall project of this thesis. The objection would be that AdS spacetime is a spacetime with negative cosmological constant, while our world is described by de Sitter spacetime, which has a positive cosmological constant. As such, any metaphysical or foundational discussion carried out in this thesis is bound to be irrelevant to the physics of our world, and hence useless.

To this objection I wish to answer, first, that it seems to me that gaining better conceptual clarity on a physical theory, even if that theory does not ultimately describe our world, is a worthwhile exercise. For example, much insight has been gained by philosophers of physics into quantum field theory (QFT) through the study of models of algebraic QFT (Haag, 1992; Halvorson, 2007), despite algebraic QFT being unable to reproduce any realistic particle physics QFT. Indeed, even though continuum QFT, which is what algebraic QFT ultimately deals with, probably is not required to describe any of the physics of our world, it has still proven useful to study these constructions in order to better understand the QFTs which do describe our world. I suggest that a similar thing is true for AdS/CFT. While AdS/CFT cannot describe our world directly, it still can provide valuable insights into how a large class of QG theories, broadly speaking those relying on holographic ideas (including ones describing positive cosmological constant spacetimes), behave. Since our world might indeed be described by one such theory, it is worthwhile to better understand conceptually AdS/CFT, as the most well-developed and under control representative of such theories.

Beyond this point, at least for the models under discussion in this thesis, it is also the case that extensions to cosmological spacetimes such as de Sitter spacetime have been developed, at least for very simple cases (Chen et al., 2021; Balasubramanian et al., 2021; Hartman et al., 2020; Bousso and Penington, 2022). Moreover, these extensions appear to preserve the basic features upon which the discussion of this thesis relies. As such, there is reason to be optimistic that the basic insights discussed here should extend straightforwardly to a holographic theory of de Sitter spacetime along those lines.

Finally, it is important to keep in mind that, as mentioned at the beginning of this chapter, the results on which this thesis relies can in principle be derived from a semiclassical analysis of the gravitational path integral, and as such do not in principle require any specific input from AdS/CFT. They are ultimately results in pure gravity and QG, even though AdS/CFT appears to be the most natural way to express them and arrive at them. As such, there is no reason to think that there is a fundamental restriction to AdS spacetimes for the constructions discussed in this thesis. In particular, the path integral computations by which these models are constructed can in principle be performed with any boundary conditions, including hence asymptotically flat and de Sitter spacetimes.

It seems then from this brief discussion that we should be optimistic regarding the possibility of extending the insights gained in this thesis to more realistic, cosmological spacetimes. Nonetheless, I will return at various points throughout this thesis to this issue, to analyze it in more detail and with more specificity to each proposal I make.

Having said this as an introduction, let me now move to describe in slightly more detail each chapter and its contents.

1.1 The Devil in the Implicit Details: on the AMPS Paradox and its Resolution

The black hole information loss paradox has long been one of the most studied and fascinating aspects of black hole physics. In its latest incarnation, it takes the form of the firewall paradox. In this chapter, I first give a conceptually oriented presentation of the paradox, based on the notion of causal structure. I then suggest a possible strategy for its resolutions and see that the core idea behind it is that there are connections that are non-local for semiclassical physics, which have to be taken into account when studying black holes. We see how to concretely implement this strategy in some physical models connected to the ER=EPR conjecture.

This chapter is based on joint work with Marco Sanchioni (Cinti and Sanchioni, 2021a).

1.2 On the Complexity of Evaporating Black Hole Spacetimes

In this chapter, my goal will be to gain some insight regarding the relevance of the global structure of spacetime for QG by looking at the semiclassical treatment of black hole evaporation and the possibility of a unitary description of black hole dynamics. In particular, my ultimate goal will be to extract some lessons from this case for two broader issues: the relevance of arguments in semiclassical gravity for QG and the peculiar role of the global structure of spacetime in QG. I do this by looking at recent results on the causal structure of evaporating black holes and how they impact various semiclassical computations of the Page curve of an evaporating black hole. Concerning both points, my conclusions point to the need for careful consideration of the theoretical context in which appeals to semiclassical gravity are made because, in QG, it seems reasonable to expect various global properties to evolve dynamically.

1.3 Amplitudes And Humeanism In QG

Humeanism has a long history of interaction with physics, especially quantum mechanics. Interestingly, recent work in philosophy of QG suggests that there might be further issues for Humeanism. In particular, this has to do with the alleged disappearance of spacetime in QG, which calls into question the usual spatiotemporal definition of the Humean supervenience base. In this chapter, I suggest an extension of Humeanism capable of dealing with these quantum gravitational worries, and illustrate its application with a concrete example from AdS/CFT.

1.4 The Fate of Spacetime In Holography

The problem of the disappearance of spacetime has long been recognized as one of the most pressing philosophical and conceptual issues facing theories of QG. In this chapter, I will look at this issue within AdS/CFT, focusing in particular on the relationship between the non-perturbative definition of QG given by the duality and the semiclassical description of gravity given by the effective field theory in the bulk. By thinking in particular about the interior of black holes and the reconstruction map connecting their effective description to their fundamental one, I come to the surprising conclusion that the standard answer to the problem of the disappearance of spacetime, i.e., emergence, is inadequate in this case, at least as usually formulated. Instead, I suggest that a more flexible and less ontologically demanding approach is required, whose basic tenet is that only operational data is required to make sense of the appearance of spacetime.

1.5 General Relativity as a Special Science

This chapter explores the status of GR as a special science when seen from the point of view of QG. In particular, it studies this issue from the perspective of the metaphysics of laws. It argues that, for certain QG cases involving a notion called operational recovery, introduced in chapter 5, only Humean approaches to laws can account for GR as a special science involving laws, while more modally committed approaches are a non-starter even in principle.

Chapter 2

The Devil in the Implicit Details: on the AMPS Paradox and its Resolution

I start in this first chapter with a discussion of the basic conceptual structure behind recent work on quantum black holes purporting to solve the firewall paradox, one of the main paradoxes in the modern literature on quantum black holes.

This topic is of fundamental importance to the philosophy of black holes, as testified by works such as Belot et al. (1999); Wallace (2020). Indeed, since Hawking (1976)'s introduction of the black hole information paradox (BHIP), and Page (1993a,b)'s subsequent development of (the basic insights behind) the Page time paradox (PTP), the topic of black hole paradoxes has been a source of deep conceptual puzzles for physicists and philosophers alike. In recent years, many essential developments came out of discussions in the high-energy physics/string theory community, centering around what is arguably the latest incarnation of the information paradox: the firewall paradox. By firewall paradox, I mean a family of arguments allegedly showing the incompatibility of the effective field theory of gravity and unitarity in the black hole interior. The first argument of this kind, and the one that will occupy us in this chapter, was *the AMPS paradox* of Almheiri et al. (2013), arguing that a unitary and semiclassical description of the black hole interior is incompatible with the monogamy of entanglement. This paradox has the interesting feature that, in contrast with its predecessors, which were largely concerned with the dynamics of evaporating black holes, here the source of the problem for the physics of quantum black holes stems from the possibility of describing the black hole interior in a way compatible with unitarity. As such, the problem is not unitary dynamics per se, or better, the paradoxes of combining unitary dynamics with semiclassical reasoning about the black hole's evolution, as in classical discussions of BHIP and PTP. Rather, the issue lies in the incompatibility of unitary dynamics with the idea that the black hole has an interior described in some approximation by semiclassical gravity. Hence, it is the black hole interior, rather than its dynamics, which takes center stage in this discussion, a theme to which I will return throughout this thesis.

Going back to this chapter, my goal will be to study the conceptual foundations of the firewall paradox and to explore how dropping an implicit assumption on the structure of spacetime, what I call **spacetime distinctness**, resolves it. In particular, I highlight, by looking at concrete physical examples, how recent discussions in AdS/CFT regarding the resolution of the firewall paradox (Papadodimas and Raju, 2013; Maldacena and Susskind, 2013; Papadodimas and Raju, 2016; Hayden and Penington, 2019; Almheiri et al., 2019; Penington, 2020; Almheiri et al., 2020) appear to rely crucially on this strategy.

To clarify the AMPS paradox's conceptual structure and its resolution, I rely on the notion of *causal structures*, which I develop in §2.1. The upshot of my analysis is that the firewall paradox crucially depends on the assumption that relativistic locality, i.e., that only causal curves can carry causal influences, broadly understood in terms of counterfactually robust correlations, is preserved in QG and that a natural way to resolve the paradox is to drop this assumption.

The chapter is structured as follows: in §2.1, I introduce the basic notion that I use throughout the rest of the chapter to extract the firewall paradox's conceptual content: causal structures. In §2.2, I give a conceptually oriented introduction to the firewall paradox and study it in terms of causal structures. In §2.3, I study some concrete physical models connected to the ER=EPR conjecture and at the core of recent discussions of the firewall paradox, and clarify that they solve the paradox precisely by dropping **spacetime distinctness**. §2.4 then concludes.

2.1 Causal Structures

I begin by introducing the notion that I employ throughout this chapter to analyze the firewall paradox:

(CS) **Causal Structure**: given a theory T, we say that the causal structure according to the theory T is given by a set of spacetime regions/objects (with their physical state) and a relation R which determines if two objects/regions of spacetime can or cannot be causally related.

To start, let me observe that when I talk about causal connections, this is done mostly for ease of exposition. By causality, I only mean that there are robust counterfactual connections between entities. No more robust notion of causality is assumed. The reader who prefers a stronger notion of causality is free to substitute for my talk of causality talk of robust counterfactual connections.[1]

To get a feel for this notion, let me use it in the case of GR. Here, the theory T is just GR, and the objects of the theory are spacetime points.[2] To capture the causal structure of the theory, then, I can define a relation R_{LC} of *being connectable by a causal curve*. This relation obtains between two spacetime points p and q if and only if a causal curve can connect them. Observe that this relation captures GR's locality properties, since it entails that only

[1]Note however that, for this discussion, one should use *robust counterfactual connections* when defining causal structures; otherwise entanglement would not fall under this notion.

[2]Note that it is not apparent how to interpret GR. For ease of exposition, I speak explicitly of spacetime points. Nonetheless, it should be possible to carry over this discussion for more refined approaches to GR's ontology.

spacetime points connectable by a causal curve can be in causal contact. Moreover, entanglement relations, since they define robust counterfactual connections, as observed in Maudlin (1994), also define a causal structure. In this case, the background theory would be QM or QFT, and the objects would be quantum systems. The relation defining the causal structure would then be R_E, i.e., *being entangled with*, obtaining between two quantum systems A and B if and only if they are entangled.

2.2 Black Hole Paradoxes[3]

In this section, I start by briefly reviewing various paradoxes regarding black holes to provide background for the subsequent discussion of firewalls. In particular, I follow Wallace (2020) in distinguishing between two main paradoxes: BHIP (§2.2.1) and PTP (§2.2.2). Then I introduce the firewall paradox (§2.2.3) and discuss its interpretation in terms of causal structures (§2.2.4).

2.2.1 Black Hole Information Paradox

BHIP concerns the non-unitarity of the complete evaporation of a black hole.[4] Consider an evaporating black hole, as in Figure 2.1. We can distinguish three regions in the diagram: (I) the pre-formation region, i.e., the collapse of a star into a black hole; (II) the evaporation region, i.e., the region in which

[3] For background material on black hole physics see Harlow (2016).

[4] See Belot et al. (1999).

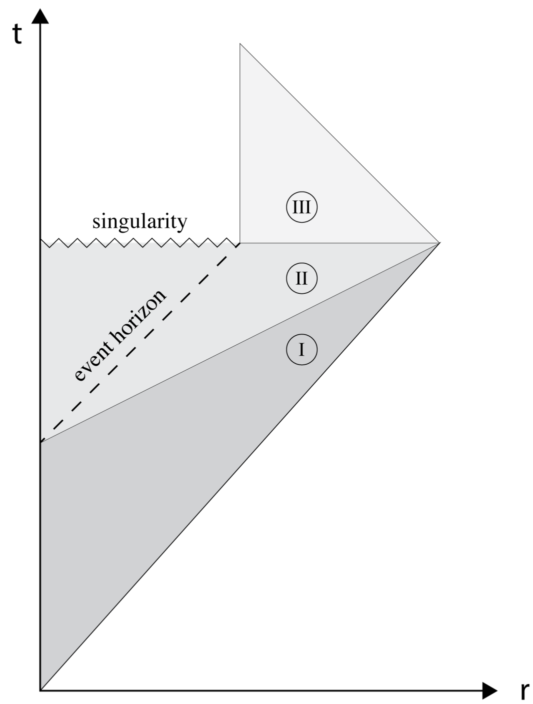

Figure 2.1: Spacetime diagram of an evaporating black hole.

the black hole starts evaporating due to the emission of Hawking quanta; (III) the post-evaporation region, i.e., the region in which the black hole has evaporated. Hawking showed that the quantum state outside the black hole in the region (II) is in a mixed state, i.e., that the state is perfectly thermal (Hawking, 1976). In other words, the quantum state of the Hawking radiation is determined only by macroscopic properties of the evaporating black hole, i.e., its mass, charge and angular momentum. In turn, Hawking radiation being exactly thermal also means that we cannot retrodict the state of the star collapsing into the black hole. The information encoded in this state should then be stored elsewhere, and the only possibility is the interior of the black hole.[5] Nevertheless, since the black hole eventually completely evaporates, this is only possible in region (II). Indeed, there are no slices in region (III) from which we can retrodict the state of the collapsing star, and thus there cannot be unitary evolution from the region (I)+(II) to region (III). It is a controversial issue whether one should take BHIP as an argument for the non-unitarity of black hole evaporation (Maudlin, 2017; Unruh and Wald, 1995, 2017) or as a true paradox. However, I do not take a specific stance on this issue, since nothing in this chapter depends upon it.

2.2.2 Page Time Paradox

Page presents a different paradox which applies long before the evaporation time (Page, 1993b).[6] Consider the black hole to be a thermodynamic system in a pure state, which we describe via a microcanonical ensemble, i.e., an ensemble whose (thermodynamic) entropy S_{MC}, where MC stands for microcanonical, is given by:

$$S_{MC}\left(E(t)\right) = \log \dim \mathcal{H}\left[E(t)\right], \tag{2.1}$$

[5] Understanding the meaning of information in BHIP is certainly important. However, I bracket this issue since nothing of what we are going to say hinges on it.

[6] See Wallace (2020) for a detailed philosophical discussion.

where $\mathcal{H}\left[E(t)\right]$ is the Hilbert space of the system at some time t and energy $E(t)$.[7] Moreover, given a composite quantum system, define the von Neumann entropy[8] of one of its subsystems as

$$S_{VN} = -\mathrm{Tr}\left(\rho \log \rho\right) \tag{2.2}$$

where ρ is the reduced density matrix of that subsystem.

Since the black hole was in a pure state before evaporation, then by unitarity at any time t after the evaporation started the composite system of the black hole and the Hawking radiation should be in a pure state. If one thinks of the radiation and the black hole as subsystems, then their Von Neumann entropies give the amount of entanglement between them and must be the same. I thus simply speak of S_{VN}. In particular, S_{VN} increases with the emission of Hawking quanta entangled with the interior. Moreover, assume that each Hawking quantum is entangled with an interior mode, to maintain the black hole plus radiation system in a pure state. Therefore, if black hole evaporation is unitary, S_{MC} bounds S_{VN}, since S_{MC} is proportional to the dimension of the Hilbert space of the black hole (2.1), and there cannot be more interior modes than the dimensionality of the black hole's Hilbert space. We call this bound the *Page bound*:

$$S_{VN} \leq S_{MC}. \tag{2.3}$$

[7]For a review of quantum statistical mechanics see Mussardo (2010).

[8]Also known as *fine-grained* entropy or *entanglement* entropy. I use these terms interchangeably in what follows.

The black hole cools down through evaporation and loses energy emitting Hawking quanta in a perfectly mixed state, which means that the microcanonical entropy S_{MC} decreases over time. On the other hand, S_{VN} increases with the number of emitted Hawking quanta, since each Hawking quanta is entangled with an interior mode.[9] Therefore, there must be a time t_P, called the Page time, at which the bound (2.3) saturates, which means that all interior modes are entangled with a Hawking quantum. The Page time t_P is also when the microcanonical entropy of the radiation becomes bigger than the microcanonical entropy of the black hole. The resulting curve, which initially grows with S_{VN} and then decreases with S_{MC}, is the *Page curve* (see Figure 2.2).

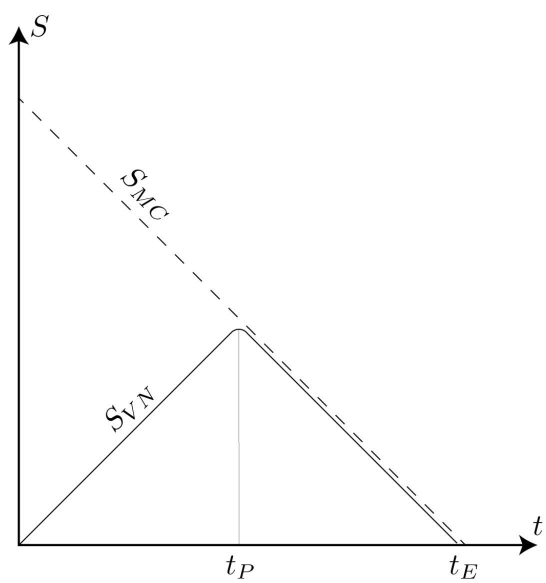

Figure 2.2: The Page curve for an evaporating black hole.

[9]Moreover, also the microcanonical entropy of the Hawking radiation increases with time since its microcanonical entropy is proportional to the number of Hawking quanta.

Consequently, since after t_P there cannot be enough interior modes to keep the composite system of black hole and radiation in a pure state, one would expect the violation of the bound (2.3) and the non-unitarity of the evaporation process. The way to avoid this conclusion is for the late-time Hawking radiation[10] not to be entangled with an interior mode, but with something else. The only possibility is that it is entangled with the *early-time radiation*, i.e., the Hawking radiation emitted at times $t < t_P$. In this way, the early radiation purifies the late, keeping the state of the black hole plus radiation system pure. Moreover, the entanglement entropy of the Hawking radiation decreases after t_P, respecting the bound (2.3). The entanglement between early and late radiation implies that the Hawking radiation is not perfectly thermal, since there are non-trivial correlations among its constituents. Nevertheless, Hawking showed that the radiation *is* perfectly thermal. Therefore, we have a contradiction. This is the nature of PTP: the inconsistency between the prediction of *naive*[11] semiclassical gravity (Hawking's calculation) and black hole statistical mechanics (the entropy bound (2.3)). PTP occurs long before the evaporation time. Indeed, it occurs at the Page time t_p, approximately half of the evaporation time for a Schwarzschild black hole.

A resolution of PTP is offered by the AdS/CFT correspondence, within which one can show that the prediction of black hole statistical mechanics is the

[10] In what follows the words *late* and *early* refer respectively to radiation emitted *after* and *before* the Page time t_P.

[11] Here by *naive* I simply mean the picture of quantum fields living on the smooth Lorentzian manifold of GR, as used in Hawking (1976).

correct one.[12] Thus black hole entropy follows the Page curve, grows with time until t_P and decreases afterward, and the bound (2.3) is not violated.

2.2.3 The Firewall Paradox

While the discussion around PTP has centered around arguments in favor or against the unitarity of black hole physics, said (non) unitarity is not exhaustive of the range of possible black hole paradoxes. In particular, in this section, I describe the firewall paradox developed in Almheiri et al. (2013) (henceforth I refer to the authors of this chapter as AMPS), which threatens the possibility of constructing a consistent theory of the interior. To start, let me follow AMPS in taking the following four postulates (originally formulated in Susskind et al. (1993) and widely accepted among high-energy physicists[13]) to be reasonable assumptions that any theory of black holes should satisfy:

Postulate 1 The process of formation and evaporation of a black hole, as viewed by a distant observer, can be described by a unitary S-matrix encoding the evolution from infalling matter to outgoing Hawking-like radiation.

Postulate 2 Outside the stretched horizon[14] of a massive black hole, physics can

[12]For a defense of this conclusion, see Wallace (2020). In particular, assuming that the correspondence is valid, black hole physics can be shown to be unitary, implying that Hawking radiation cannot be thermal.

[13]At least among physicists in the string theory community. Note that these postulates were originally defined in the context of black hole complementarity. However, they are not strictly tied to this specific proposal. Rather, they are taken to be definitive of what a sensible theory of black hole physics should look like.

[14]An horizon at a distance of a Planck length from the true event horizon. It is time-like.

be described to a good approximation by a set of semiclassical field equations.

Postulate 3 To a distant observer, a black hole appears to be a quantum system with discrete energy levels. The dimension of the subspace of states describing a black hole of mass M is the exponential of the Bekenstein entropy S_{MC}.

Postulate 4 A freely falling observer experiences nothing out of the ordinary when crossing the horizon until the singularity is approached. Another way to say this is that no observer ever detects a violation of the known laws of physics.

AMPS showed that these four postulates are inconsistent. Indeed, take an evaporating black hole, fix a spatial hypersurface at time $t > t_p$, and consider its Hawking radiation. Divide the black hole plus radiation system into three subsystems: early radiation E, late radiation L, and the interior partners to the late radiation B. As we have seen in the previous section, the fact that the Hawking radiation is in a pure state means that the late radiation L and the early radiation E should be entangled. Indeed, a subsystem of a bipartite system in a pure state is maximally entangled with its counterpart. Therefore, since the combined system of early and late radiation is pure (a consequence of Postulate 1), then the early radiation and the late radiation should be maximally entangled. Moreover, if there is no drama at the horizon (Postulate 4), the late radiation L is fully entangled with the modes behind the horizon B. Indeed, for an infalling observer, the geometry near the horizon can be locally identified with a Rindler horizon via a coordinate transformation:

$$\tau = \frac{t}{4M} \quad , \quad \rho = 2\sqrt{2m(r-2M)} \quad , \quad \tilde{x} = (2M\theta, 2M\phi)\,, \tag{2.4}$$

which turns the Schwarzschild metric into

$$ds^2 \approx -\rho^2 d\tau^2 + d\rho^2 + d\tilde{x}^2, \tag{2.5}$$

the metric corresponding to Rindler spacetime.[15] Fields in the left Rindler wedge are maximally entangled with fields in the right Rindler wedge. Therefore, a mode L outside the horizon must be maximally entangled with a mode B inside the horizon. Thus, the composite system, made of E, L and B, has two features: L is maximally entangled with both E and B.

However, a fundamental property of entangled systems is what in quantum information is called the monogamy of entanglement: a quantum system can be maximally entangled with *only one* quantum system at a time. Equivalently, monogamy of entanglement says that there is an upper bound on the independent degrees of freedom with which a given quantum system can be entangled, given by the number of independent degrees of freedom of the quantum system itself. Thus, the four postulates are in contradiction with the monogamy of entanglement. This contradiction is known as the firewall

[15] Note that, here, $d\tilde{x}^2 = d\tilde{x}_1^2 + d\tilde{x}_2^2$, where $x_1 = 2M\theta$ and $x_2 = 2M\phi$.

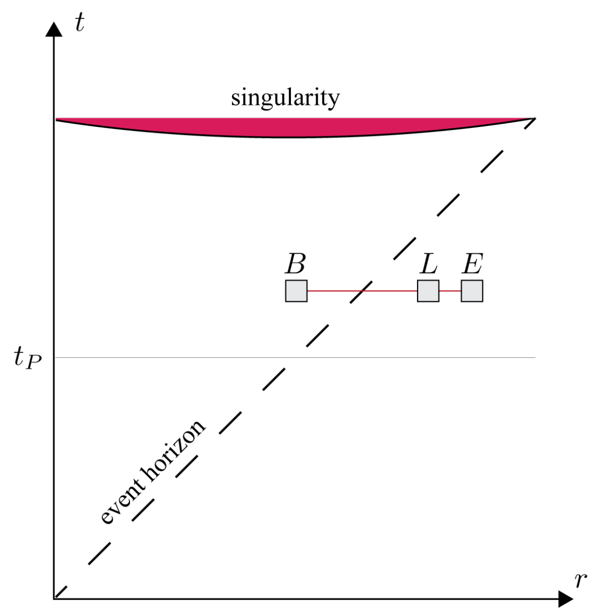

Figure 2.3: A sketch of the basic scenario behind the firewall paradox. B is an interior mode, L is a mode of late Hawking radiation, while E is a mode of early Hawking radiation. The red lines represent relations of maximal entanglement.

paradox.[16] AMPS proposed that L and B are not entangled, but are instead in a product state. Since L and B being in a product state means that close to the horizon Rindler space is not a good model for spacetime, the assumption that an observer does not see anything out of the ordinary there [Postulate 4] is not reliable. Indeed, if there is no entanglement between L and B, B is not entangled with anything. On the other hand, L is still entangled with E, which purifies it. Therefore, the state of B is characterized by its reduced density matrix, which is thermal.[17] The thermal nature of B leads to a high concentration of energy at the horizon, which can be suggestively described as a wall of fire, hence the notion of the firewall.

The existence of firewalls implies that a non-trivial theory of the interior is impossible since the interior would always be characterized by the same thermal state, i.e., the firewall. As such, no interesting theory of the black hole interior would be possible in the presence of firewalls.

2.2.4 AMPS and Causal Structures

Let me now clarify where the tension lies in the firewall argument of AMPS. The standard formulation of the paradox is useful in showing the incompatibility of the four postulates. What remains unclear is where precisely the

[16] An equivalent formulation of the paradox is as a violation of strong subadditivity of entanglement entropy, i.e. $S_{AB} + S_{BC} \geq S_B + S_{ABC}$.

[17] The thermal nature of the state of B is because, since we want the entangled and non-entangled descriptions to be locally indistinguishable, to describe the state of B we use the reduced density matrix of the maximally entangled state of the composite system of B and L. It is a well-known fact that the reduced density matrix of a subsystem of a maximally entangled system is completely thermal.

conflict giving rise to the incompatibility between relativity and quantum theory lies. To make clear what is the conceptual origin of this conflict, I make use of the notion of causal structure.

Beyond the four postulates listed, the fact that B and E are separate systems plays a fundamental role in the paradox. For black holes, the distinctness between B and E is justified because B lies behind the horizon of the black hole. B is thus causally isolated from L and E since, within the fixed spatial hypersurface at time $t > t_p$ used to formulate the AMPS paradox, the black hole's interior and exterior are spacelike separated.[18] However, this is not an assumption as innocent as it might sound: it is equivalent to the claim that relativistic notions of separability and locality, based on spacelike and causal connections, are retained in the regime of QG. Let me focus on this point, as it will be crucial in the rest of this chapter.

As I have mentioned, to ensure the violation of monogamy, and thus the firewall paradox, one has to regard B and E as distinct systems, where by distinct I mean that their degrees of freedom are independent. This fact follows from monogamy of entanglement giving an upper bound on the number of independent degrees of freedom with which a certain quantum system, in our case L, can be entangled. Indeed, (i) we know that B and E are spacelike related since, at fixed t, one is in the interior and the other in the

[18]Even if one were to consider an observer who takes E and L and jumps into the black hole meeting B, thus making E and B not spacelike separated, it is still the case that the distinctness of E and B, and thus the AMPS paradox, is grounded in their being spacelike separated before the observer jumping in. Thus, in what follows, I will not consider this situation and stick to the more general formulation of the paradox outlined above.

exterior of the black hole. Furthermore, (ii) from the study of QFT (and in particular of Algebraic QFT[19]) we have come to accept that if two algebras of observables[20] are mutually commutative, then they represent two distinct systems,[21] since, when this is the case, their degrees of freedom are completely independent. From the axiom of *microcausality*, (iii) we have that the algebras of observables connected to two spacelike related regions must commute, i.e., $[\mathfrak{N}(\mathcal{X}), \mathfrak{N}(\mathcal{X}')] = 0$, where $\mathfrak{N}(\mathcal{X})$ and $\mathfrak{N}(\mathcal{X}')$ are the algebras of observables associated to two spacelike related regions $\mathcal{X}$ and $\mathcal{X}'$. Thus, from (i)-(ii)-(iii), the two systems must be distinct.

Note however that the microcausality axiom encodes the locality properties of classical relativistic spacetime since it relies explicitly on the notion of spacelike separation, which we have no guarantee will be retained at the level of quantum spacetime. Thus, we have to assume that this notion of locality, developed for relativistic spacetimes, can be extended seamlessly beyond GR. If this were not the case, then B and E might not be distinct at the quantum level, opening the door to resolving the paradox by observing that there is no violation of monogamy since B and E are not distinct systems which implies that their degrees of freedom are not independent and that L is entangled with fewer degrees of freedom than those manifest in the semiclassical description (more on this in §2.3).

We have thus seen that to the four postulates above, AMPS (implicitly) add

[19]For reviews of AQFT see Halvorson (2007); Haag (1992).

[20]Note that its algebra of observables identifies a system.

[21]Though see Earman and Valente (2014) for some subtleties about this claim.

a fifth one:

(SD) Spacetime Distinctness: spacelike separated systems are distinct, i.e., mutually commuting.

With this in mind, we can then reformulate the paradoxical conclusion of AMPS as follows: in a black hole spacetime, we have three subsystems L, B and E, such that B and L are distinct systems, i.e., $\mathfrak{N}(B)$ commutes with $\mathfrak{N}(L)$, and they violate the monogamy principle, i.e., L is maximally entangled with both B and E.

To better keep track of the various moving parts of the AMPS argument, and to elucidate its conceptual content, let me recall the definition of causal structure:

(CS) Causal Structure: given a theory T, we say that the causal structure according to the theory T is given by a set of spacetime regions/objects (with their physical state) and a relation R which determines if two objects/regions of spacetime can or cannot be causally related.

To recast the AMPS paradox in terms of causal structures, let me take as our objects quantum systems and let me consider two different relations, which define two different causal structures. One is the relation R_{LC} of *being connectable by a causal curve* that we have encountered in §2.1, which defines what I call the causal structure of spacetime. The other is the relation R_{ME} of *being maximally entangled*, which defines what I call the causal structure of entanglement, and is a slight restriction of the relation R_E seen in §2.1.

The core claim of AMPS is then that, under the four postulates detailed in §2.2.3, there are three quantum systems L, B and E such that $R_{ME}(L, B)$ and $R_{ME}(L, E)$, and B and E are distinct, which follows from $\neg R_{LC}(B, E)$, i.e., B and E are spacelike related. Our four assumptions (supplemented with **spacetime distinctness**) are then in violation of the monogamy of entanglement (see Figure 2.3). In other words, combining the causal structure of spacetime and entanglement by simply superimposing them leads to a contradiction. To resolve this paradox, one has three possibilities:

(i) accept $\neg R_{ME}(L, B)$, i.e., L and B are not entangled. This is AMPS' answer and implies a firewall at the horizon.

(ii) accept $\neg R_{ME}(L, E)$, i.e., L and E are not entangled. This answer is equivalent to Hawking's calculation and implies the non-unitarity of black hole evaporation.

(iii) accept that B and E are not distinct. This answer implies, as we see in §2.3, the modification of the causal structure of spacetime.

The solution that I study in this chapter is (iii). Let me briefly remark why I think it is more promising than (i) and (ii). First, (i) entails the violation of the equivalence principle of GR, since it implies that freely falling observers do not experience the gravitational vacuum at the horizon but instead meet the firewall. Since the equivalence principle is one of the fundamental insights of GR, one should be careful about renouncing it. As regards (ii), there is a violation of unitarity. However, the non-unitarity in Hawking's calculation comes from Plank scale effects, which take place at the end of the evaporation process. On the other hand, the AMPS paradox occurs long before this time, when semiclassical gravity and effective field theory are still approximately valid descriptions. As such, it is unclear how this violation of unitarity might arise in this context. Moreover, as we have observed at the end of §2.2.2, unitarity seems to be a core aspect of quantum theory and to be retained in QG, at least insofar as one can trust arguments based on the AdS/CFT correspondence. As such, both (i) and (ii) seem to imply the violation of well justified, though not immune from revision, principles. On the other hand, (iii) would imply, as I remarked above, that the relativistic notion of locality does not apply to quantum spacetime. However, the justification for thinking this seems to be shakier than the other principles discussed. To my knowledge, there is no explicit argument in defense of this claim. As such, I take it that before accepting (i) or (ii), one should at least test the viability of (iii), as it requires milder revisions to the fundamental principles of physics than its alternatives.

2.3 No Firewall on the Horizon

In this section I am going to introduce some concrete physical models, constructed in the context of AdS/CFT and of the ER=EPR conjecture (Maldacena and Susskind, 2013).[22] These models play a crucial role in recent

[22]Here, ER stands for Einstein and Rosen, from the seminal article Einstein and Rosen (1935) introducing wormholes, while EPR stands for Einstein, Podolsky, and Rosen from

discussions of the firewall paradox within the high-energy physics community (Papadodimas and Raju, 2013; Maldacena and Susskind, 2013; Papadodimas and Raju, 2016; Hayden and Penington, 2019; Almheiri et al., 2019; Penington, 2020; Almheiri et al., 2020).

The point of looking at these models is to understand how they avoid falling into the pitfalls of the AMPS paradox. Indeed, my goal in this section will be to argue that these constructions attack the AMPS paradox by following route (iii) and dropping **(SD)**. We will see that the non-local connections characteristic of the ER=EPR conjecture, and concretely instantiated in these models, engineer the violation of **(SD)** in which we are interested. Thus, how ER=EPR avoids the firewall paradox is a paradigmatic instance of route (iii)'s strategy for resolving the paradox.

I proceed in §2.3.1 by discussing the instructive, though not realistic, case of the eternal AdS black hole where we can see how ER=EPR undermines **(SD)** fundamental to the AMPS paradox. In §2.3.2, I then move to the case directly relevant to the AMPS paradox, that of an evaporating black hole formed from gravitational collapse. Finally, in §2.3.3, I discuss how these ideas relate to another important approach to describing the black hole interior, black hole complementarity.

---

the article Einstein et al. (1935) which first pointed out the EPR paradox and the non-local character of entanglement.

2.3.1 Eternal Black Holes

Let me start from the case of eternal[23] AdS black holes, following the treatment of Maldacena and Susskind (2013). While these are not the evaporating black holes generated from gravitational collapse, for which the AMPS paradox applies, they can still be instrumental in testing and developing ideas regarding black holes' structure. In particular, the purpose of starting from the case of a two-sided black hole[24] is that it serves a useful pedagogical role since it allows to phrase various questions regarding the interior structure of black holes to a level of precision hard to attain in the context of standard, evaporating, one-sided black holes.[25] It is thus useful to start from this most basic case.

Eternal AdS black holes are holographically dual to a couple of entangled CFTs (called the *left* and *right* CFT depending on which of the two exterior regions of the AdS black hole they describe) in the so-called *thermofield*

[23]By eternal here I mean a black hole which has always existed and will always exist. Thus, the black hole has not formed via gravitational collapse, and also it is not subject to evaporation. In AdS, the non-evaporation of the black hole is due to the reflecting boundary conditions of the spacetime, which mean that the Hawking radiation emitted from the black hole bounces back inside the black hole upon reaching the boundary of AdS spacetime. Thus, the black hole is in equilibrium with exterior spacetime and does not evaporate. To make an AdS black hole evaporate, one needs to have absorbing boundary conditions, which I discuss in §2.3.2.

[24]By a two-sided black hole I mean a physical system with two event horizons. As such, a two-sided black hole has two exterior regions, usually called the *left* and the *right* exterior. Eternal AdS black holes are two-sided black holes.

[25]By a one-sided black hole I mean a physical system with one event horizon. As such, a one-sided black hole has only one exterior region.

double state:

$$|\psi\rangle = \sum_j e^{-\beta E_j/2} |j\rangle_L \otimes |j\rangle_R \tag{2.6}$$

where E_j are the energy levels of the CFT and $\beta = 1/T$, where T is the temperature of the black hole, $|j\rangle_L$ and $|j\rangle_R$ are states in the left and right CFT. The core of ER=EPR is the conjectured equivalence between entangled systems and wormhole geometries, i.e., that between any two entangled quantum systems there is a wormhole, possibly of Planckian size.[26] The eternal AdS black hole has a dual interpretation: we can either understand it as a system made of two entangled black holes or as two black holes connected by a wormhole. This dual way of looking at eternal AdS black holes is at the heart of the ER=EPR proposal. Indeed, the eternal AdS black hole is a particularly special case of the ER=EPR conjecture, in the sense that we do not need to modify the classical geometry of the black hole to get ER=EPR. As it were, the wormhole is already there in the eternal AdS black hole. In particular, no quantum wormhole is needed to verify the conjecture in this case, only classical geometry, making the subsequent discussion much easier. The basic lessons that we learn in this case, however, carry over also to the more realistic cases that I treat later, where no such convenient semiclassical picture is available.

[26] Susskind (2016a) distinguishes between a modest and an ambitious version of the conjecture. Modest ER=EPR is supposed to apply only to entangled black holes, while ambitious ER=EPR applies to any entangled systems. In this chapter, I am concerned only with the modest version, since I only talk about black holes.

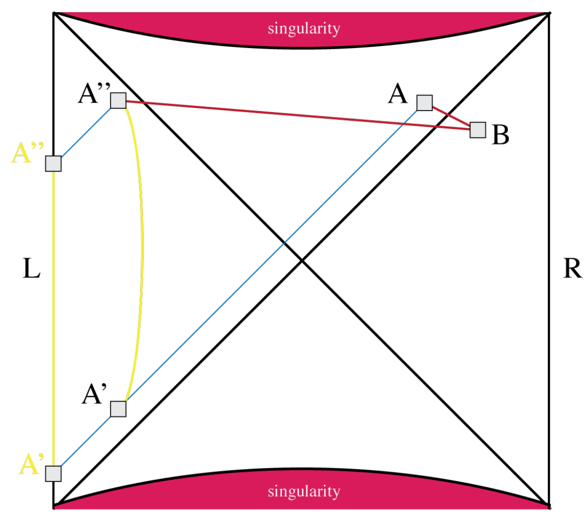

Figure 2.4: The Penrose diagram of an eternal AdS black hole (with L the left exterior and R the right exterior) with an AMPS situation and its ER=EPR resolution. Here, red lines represent entanglement and blue lines semiclassical bulk evolution, while yellow lines represent boundary evolution and its bulk dual. Furthermore, A is an interior mode, A' is its CPT transform (with its holographic dual in yellow), and A'' is a mode on the left horizon (with its holographic dual in yellow) obtained by evolving A'. The AMPS-like situation comes from the maximal entanglement between A'' and B, and A and B. The resolution is given by the existence of a unitary connection between A'' and A.

The semiclassical features of the eternal AdS black hole are at the heart of how ER=EPR solves the AMPS paradox. To see why, let me start by constructing a firewall like situation in the context of the eternal AdS black hole (Figure 2.4). We start at time t by taking a pair of entangled qubits A

and B, on the right side of the two-sided eternal black hole, with A behind the black hole horizon and B outside the horizon. Let us now apply a CPT transformation[27] to A. Since this transformation is a symmetry, sends t into $-t$, and exchanges left and right, we get a qubit A', equivalent to A, at time $-t$ in the left exterior region of the black hole. Furthermore, it is clear from the Penrose diagram, that we can evolve A' into A with the bulk equations of motion. Let us write this as $A' \to A$. In particular, A' is entangled with B since we just applied a CPT transformation to A, and A is entangled with B. Since A is a qubit, it has three components A_i where $i = 1, 2, 3$, which are Pauli matrices. We thus have that:

$$[A_i, A_j] = i\epsilon_{ijk}A_k \neq 0 \tag{2.7}$$

Since, however, by forwards time evolution with the bulk equations of motion, we have that A' evolves in A, we can also write:

$$[A'_i, A_j] \neq 0 \tag{2.8}$$

To arrive at this result, we rely on the fact that we can view the eternal AdS black hole as a wormhole connecting two horizons. A crucial step was the forwards time evolution of A' in A, relying on A' passing through the wormhole, as evident in Figure 2.4. However, we also know that we can re-

[27]CPT here stands for *charge*, *parity* and *time reversal*. CPT transformations are among the fundamental symmetries of quantum systems, such as the one we are considering here.

gard the eternal AdS black hole as two entangled black holes, which means that we can evolve qubits in the left exterior (dual to the left CFT) independently of the right exterior (dual to the right CFT) since these are just two disconnected spacetimes. In particular, if we evolve A' forwards in time up to time t, with the left CFT Hamiltonian, we obtain the qubit A'', which is naturally understood as the holographic dual to a qubit living on the left horizon.[28] Since A'' is just the product of evolving forwards in time A', it carries the same information (in the sense that they are related by a unitary operator, the left CFT Hamiltonian). In particular, it is entangled with B, giving us a firewall like situation. At t, we have a qubit B entangled with a qubit in the interior (A) and a distant, far away qubit (A''). Here B is equivalent to L in my formulation of AMPS while A is equivalent to B, and A'' to E (remember Figure 2.3). It would thus seem that here too we have a violation of the monogamy of entanglement. However, we can immediately see the resolution of this apparent paradox. Since A'' follows from applying the left CFT Hamiltonian to A', and $A' \rightarrow A$, we can also write $A'' \rightarrow A$.[29] But then we can substitute A'' for A' in (2.8), giving us:

$$[A_i'', A_j] \neq 0 \tag{2.9}$$

[28]To be precise, the left stretched horizon, though this is not relevant to the present discussion.

[29]Intuitively, we can understand this as first evolving backwards in time from A'' to A' with the left CFT Hamiltonian, and then forwards in time from A' to A with the bulk equations of motion.

(2.9) resolves the apparent violation of monogamy that we have engineered since it tells us that A and A'' are not independent qubits (since they do not commute). As such, it is not the case that B is maximally entangled with two different systems, leading to a violation of the monogamy of entanglement. Since A and A'' are not distinct systems, B is entangled with only one system, represented by both A and A'', in a fundamentally non-local manner. The mistake in the reasoning which led us to an apparent violation of monogamy was the (background) assumption of **(SD)**, which told us that two spacelike separated systems must commute, and thus be distinct. Since A and A'' are spacelike separated, it was only natural to assume that they were two different, distinct systems. What my analysis shows is that, in the context of the eternal AdS black hole where ER=EPR is already a feature of the semiclassical geometry, the assumption of **(SD)** falls apart, which is the essence of (iii)'s resolution of the AMPS paradox. Even spacelike separated, distant objects can still nonetheless depend on each other, and thus not be distinct objects after all.

This analysis is best understood in the language of causal structures **(CS)**. In §2.2.4, we have seen that the monogamy paradox can be recast as the mismatch between the causal structure of spacetime and the causal structure of entanglement. Since $R_{ME}(A, B)$ and $R_{ME}(A'', B)$ while $\neg R_{LC}(A'', A)$, which, by **(SD)**, implies that A'' and A are distinct systems, we come to the conclusion that B is maximally entangled with two distinct systems, violating monogamy. The way ER=EPR allows us to resolve the paradox, in the context of the eternal AdS black hole, is by defining a more general causal structure, characterized by the relation R_{WH} of *being non trivially connected*, which obtains if and only if two entities A and B can non trivially influence each other. A non-trivial influence is manifested by the presence of counterfactually robust correlations between A and B. I call this the *generalized causal structure.*

Observe that it is always the case that we can embed the causal structure of spacetime and entanglement in the generalized causal structure, by observing that both R_{ME} and R_{LC} are supposed to produce robust counterfactual correlations. However, and here lies causal structures' usefulness in understanding the ER=EPR resolution of AMPS, in certain situations such an embedding is not an isomorphism, i.e., there are systems connected by R_{WH} which are connected neither by R_{ME} nor R_{LC}. Indeed, in the eternal black hole in AdS, the wormhole connects the left and right exterior, making it possible to have a non-trivial connection between A and A'', despite their not being entangled and being spacelike related. Indeed, it is this non-trivial connection, captured by R_{WH}, which leads to the violation of **(SD)**, as per (iii).

The eternal AdS black hole serves as a simple motivating case to understand how violations of **(SD)** naturally emerge in the study of black holes. We can now move on, in the next section, to the study of evaporating black holes formed from gravitational collapse. Let me, however, remark once more the most critical intuition underlying the ER=EPR conjecture: it is the idea that the overall structure of spacetime, as it emerges from QG, is much more complicated than its *naive* semiclassical description would lead us to believe. There are many more connections that are not accounted for by merely thinking in terms of causal curves, and these connections are central to a proper understanding of black holes. Furthermore, while in some lucky cases (such as the eternal AdS black hole) we can understand these connections in geometrical terms as wormhole geometries, in other cases this is not possible. Instead, we have to resort to the more general idea of there being connections absent in the semiclassical description, imaginatively called *planckian* or *quantum* wormholes in Maldacena and Susskind (2013). These non-local (from the perspective of semiclassical spacetime) connections are the heart of the ER=EPR conjecture and lie at the core of the constructions described here.

2.3.2 Evaporating Black Holes from Gravitational Collapse

In this section, I study the firewall paradox in a situation where the black hole is formed from gravitational collapse and can evaporate, a problem studied extensively in Penington (2020).

As I already said, the black hole in AdS spacetime considered in §2.3.1 is not evaporating, since it has reflecting boundary conditions, and cannot give a real example of the firewall paradox. Indeed, the conformal boundary of

AdS reflects the Hawking quanta into the black hole, reaching thermal equilibrium since the number of emitted and reflected quanta compensate (see Figure 2.5).

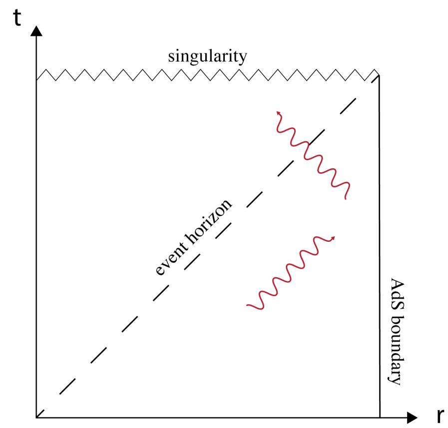

Figure 2.5: The Penrose diagram for a black hole with reflecting boundary conditions. Observe that the radiation (wiggly red lines) bounces on the boundary and comes back to the black hole.

To build an evaporating black hole, take a black hole formed from collapse and place it into a spacetime whose boundary is not completely reflecting, i.e., in a spacetime that permits some Hawking quanta to escape outside the AdS boundary. In this case, the emitted Hawking radiation will be larger than the Hawking radiation coming back into the black hole, since some radiation has escaped outside the boundary. The black hole, then, slowly evaporates. The more radiation we permit to escape from AdS, the faster the black hole evaporates. This procedure can be made precise within the context of AdS/CFT by coupling the boundary CFT to an auxiliary reservoir $\mathcal{H}_{rad}$, using absorbing boundary conditions. Furthermore, I assume that $\mathcal{H}_{rad}$

is a large holographic system, which allows the (holographic) encoding of the Hawking radiation into $\mathcal{H}_{rad}$ (see Figure 2.6). How can we incorporate H_{rad},

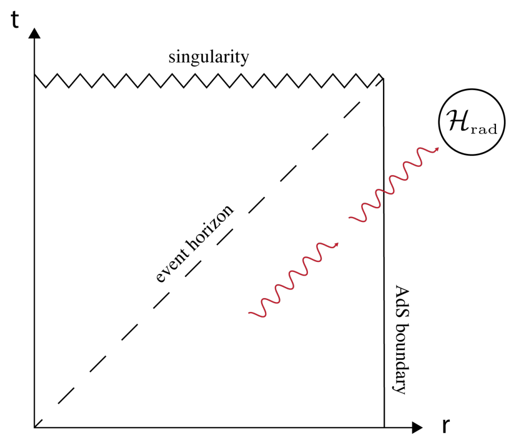

Figure 2.6: The Penrose diagram for a black hole with absorbing boundary conditions. Observe that some radiation (wiggly red lines) escapes from the boundary and is holographically encoded in $\mathcal{H}_{rad}$.

and thus the escaping Hawking radiation, into the analysis of §2.3.1? Take for instance a quantum of late Hawking radiation L, which, after the Page time, is entangled with some interior modes B and with the early Hawking radiation E, thus having a monogamy problem. As in the two-sided black hole case of §2.3.1, where the interior mode A was encoded in the left CFT, the interior mode B is encoded in the boundary theory. In particular, for one-sided evaporating black holes, the mode B is encoded in $\mathcal{H}_{rad}$.[30] Before the Page time t_p, $\mathcal{H}_{rad}$ encodes only the early Hawking radiation that escapes from the boundary of AdS. After the Page time, t_p, $\mathcal{H}_{rad}$ also encodes B, de-

[30]This statement is the core of the analysis of Penington (2020), where it is proven using entanglement wedge reconstruction.

spite it being in the interior and thus unable to reach the boundary. This fact signals the breakdown of the semiclassical picture since we have two spacelike separated systems B and E, which live in the same CFT $\mathcal{H}_{rad}$. Thus, there is no problem for the late Hawking radiation L to be entangled with both the interior mode B and $\mathcal{H}_{rad}$ (which encodes the escaped early radiation E), since, as in the case of the eternal AdS black hole, the first statement implies the second one. Since the interior mode B is encoded in $\mathcal{H}_{rad}$, entanglement with B implies entanglement with a mode in $\mathcal{H}_{rad}$. Equivalently, **(SD)** is violated since we have two spacelike separated systems, one inside and one outside the black hole, B and $\mathcal{H}_{rad}$ (which encodes E) which are nonetheless not distinct systems. Their not being distinct is a consequence of the fact that B is encoded in $\mathcal{H}_{rad}$, which equivalently means that B is a part of $\mathcal{H}_{rad}$. Thus, the two systems cannot be distinct.

From the perspective of causal structures, we can analyze this situation in the same way in which we have studied the eternal AdS black hole of §2.3.1. The paradox is that it seems to be the case that $R_{ME}(L, B)$ and $R_{ME}(L, E)$, while $\neg R_{LC}(E, B)$, which by **(SD)** implies that B and E are two distinct systems, violating monogamy. However, as we have seen in this section, although E and B are spacelike separated, they are not distinct. Indeed, after t_p, B is holographically encoded in $\mathcal{H}_{rad}$ and therefore connected with E. This new connection can be encoded via R_{WH} and corresponds to a connection which is captured neither by R_{LC} nor by R_{ME}. Again, the generalized causal structure (defined via R_{WH}) captures the structure of the black hole that goes beyond the semiclassical approximation, which relies only on R_{LC} and R_{ME}. In particular, R_{WH} encodes those connections which show that the two systems B and E which, from the perspective of the causal structure of spacetime and entanglement, are distinct and separated, are interdependent and connected, thus violating **(SD)**.

Furthermore, this is again the same intuition of ER=EPR, that the semiclassical picture of spacetime crucially fails in the context of black holes in taking into account non-local connections that are neither causal curves nor entanglement relations. One of the main advantages of the causal structure approach is that we can naturally show the underlying strategy behind the different proposals for resolving the AMPS paradox that I have studied thus far.

2.3.3 Black Hole Complementarity Regained?

In the previous sections, we have seen how the ER=EPR conjecture and related constructions were able to resolve the firewall paradox by violating **(SD)** and hence instantiating route (iii). The overall strategy behind these approaches is to show that the interior and the exterior of the black hole are not independent systems, i.e., the violation of **(SD)**. From the perspective of causal structures, there is a connection R_{WH} which is neither a spacetime relation R_{LC} nor an entanglement connection R_{ME}. Interestingly enough, the core of this strategy appears to be a precise incarnation of the basic principle of black hole complementarity (BHC) (Susskind, 2012).

BHC is the idea that the information falling towards a black hole both *passes through* and *is reflected at* the horizon. The outside and the inside of a black hole are two complementary, non-commuting, descriptions of the same physics. For an observer outside the black hole, the inside does not exist. For an observer inside the black hole, the outside does not exist. Moreover, when the infalling observer passes through the event horizon, he perceives nothing special to happen. However, according to an observer outside the black hole, the horizon heats up, burning the infalling observer. The outgoing Hawking radiation encodes the information regarding the infalling observer in the same way in which smoke and dust encode information about a book that has just been burnt. However, there is no contradiction since no observer has access to both realities. It is like the two observers, inside and outside the black hole, are viewing the same physics from two different (non-commuting) points of view.

To see the violation of **(SD)** and route (iii) relate to BHC, one should observe that the interior and the exterior not being independent systems is equivalent to the claim that they do not commute. This statement is, at the same time, the core of BHC and the reason we can recover information about the interior from the exterior and vice-versa, rendering the AMPS paradox moot. These considerations, then, point towards the fact that the AMPS paradox is not an argument against the consistency of BHC, as its authors initially believed. It is instead an argument against the consistency of BHC together with **(SD)**.

Nevertheless, **(SD)** was never a reasonable assumption for those who believed in BHC. However, BHC was missing an explicit construction of the black hole interior, finally provided by the ER=EPR conjecture and its various generalizations. As we have seen, these constructions instantiate the core intuition of BHC, i.e., the non-commutativity of the interior and the exterior. The mistake behind the AMPS construction is even more evident when one thinks of the causal structures involved in this argument. AMPS believed that the appropriate description of the interior and the exterior of the black hole should be given as the superposition of entanglement and spacetime causal structures. This construction then leads to the assumption of **(SD)** and to the firewall paradox itself. However, such a construction of the interior is not acceptable from the perspective of BHC.

Since the interior and the exterior are non-commuting descriptions of the same physics, there must be some new type of connection, which we have called R_{WH}, which makes this interdependence manifest. Furthermore, R_{WH} cannot be captured either by entanglement or by spacetime causal structures. Indeed, in general, it is not part of the semiclassical descriptions of black holes.

These connections R_{WH} are precisely the type of connections that differentiate the generalized causal structure from the spacetime and entanglement causal structures. As such, the generalized causal structure is the correct structure to encode the description of the black hole interior. As we have seen, at least in the context of holographic theories of gravity and in particular AdS/CFT, this is indeed the case. It is precisely this difference between the generalized causal structure defined by R_{WH} and the entanglement and spacetime causal structures that allows BHC to remain consistent in the face of the AMPS paradox, and thus to remain a viable candidate for a theory of the interior of the black hole.

2.4 Conclusions

In this chapter, we have seen how the AMPS paradox appears to threaten the consistency of black hole physics and how dropping the implicit assumption of **spacetime distinctness** allows us to overcome it. The core of my work is the notion of causal structure. In particular, this notion helped me in highlighting the role of **(SD)** and in making explicit the link between various strategies for constructing the interior of the black hole, as explained in §2.3.1 and §2.3.2. Furthermore, causal structures make clear the sense in which ER=EPR-like connections imply a violation of semiclassical locality.

An especially important issue that immediately emerges from this discussion is the status of spacetime in the models discussed here. Indeed, the crux of the violation of **(SD)** lies in its modifying the semiclassical structure of spacetime to include further connections. How should we think of these connections, however? And how do they change our view of the ontology of spacetime? ER=EPR suggests that we view these connections geometrically, as wormholes, and thus include them into the structure of spacetime. At the same time, the idea itself of planckian wormholes, and more generally that interior and exterior are described by the same degrees of freedom as discussed in §2.3, suggests that spacetime as usually understood might not be the correct tool to make sense of these models. These questions are of clear importance for a full understanding of the structure of holographic quantum black holes.

However, in the next chapter, rather than the ontology of spacetime, I will look at the foundations of various path integral derivations of the models and results discussed in this chapter. Insofar as these computations provide an independent derivation of these results, relying only on the reasonably well understood physics of the semiclassical approximation of the gravitational path integral, understanding them is a task of supreme importance for those interested in quantum black holes and AdS/CFT, and is moreover a natural continuation of the analysis of this chapter.

Chapter 3

On the Complexity of Evaporating Black Hole Spacetimes

As we have seen in the previous chapter, black hole evaporation, and the various paradoxes stemming from the attempt to reconcile this phenomenon with the basic rules of quantum mechanics, have played a crucial role in the development of QG since their inception. Answering these challenges has time and again provided us with important insights into the structure of QG. Not only that, but they have also been able to illustrate and put in a particularly sharp light various issues in the philosophy of physics, such as the viability of a statistical mechanical interpretation of entropy or the validity of the second law of thermodynamics.[1] Indeed, from a philosophical point of view, this is arguably the main reason for interest in black hole evaporation.

A crucial step forward, at least from the perspective of those working on these issues from the perspective of high-energy physics, has been the computation of the Page curve for an evaporating black hole from purely semiclassical considerations in the context of the AdS/CFT that I have discussed in the previous chapter in the context of the AMPS paradox and its resolution (Almheiri et al., 2019; Penington, 2020). Since the Page curve is widely accepted as a critical signature for unitary evolution, these computations are taken to provide convincing evidence for the unitarity of black hole evaporation. Even more interestingly, right after this holographic computation of the Page curve, the same models and formulas have been derived from computations via the Euclidean gravitational path integral treated semiclassically (Almheiri et al., 2020; Penington et al., 2019). Note that by a semiclassical treatment of the gravitational path integral I will mean, roughly, an evaluation of the gravitational path integral through an expansion around a classical spacetime geometry. The result of these calculations is a computation of the Page curve which does not rely on any detailed QG assumption but instead only on semiclassical considerations, which significantly increased the belief of physicists working in the field in the validity of the result: if semiclassical

[1] See Belot et al. (1999); Wallace (2020); Curiel (2019); Cinti and Sanchioni (2021a) for philosophical discussion of issues having to do with black hole evaporation.

gravity is sufficient to derive these results, then they are as general as they could ever be.

At the same time, these results have some limitations:

(i) They have been obtained explicitly only in the context of a specific two-dimensional theory of gravity known as JT Gravity (Teitelboim, 1983; Jackiw, 1985), where the path integral can be treated non-perturbatively, and thus their extension to higher dimensions, while expected, is to be checked.

(ii) They appear to be in contrast with a wealth of results in semiclassical GR which imply that black hole evaporation spacetimes violate various causality conditions (Kodama, 1979; Wald, 1984; Manchak and Weatherall, 2018; Lesourd, 2018), and are thus incompatible with the well-behaved propagation of quantum fields, hence with unitarity.[2]

These two issues are intimately related, since the existence of results like (ii) would imply the limitation that (i) is a necessary condition on these computations, and thus they cannot be extended to higher dimensions. Therefore, it seems that the generality of these results, which have been so influential in recent discussions of black hole paradoxes, is in danger.

This chapter aims to explain why the semiclassical computations of the Page curve of Almheiri et al. (2020); Penington et al. (2019) avoid contradiction

[2]Note that these results do not explicitly refer to Penington et al. (2019); Almheiri et al. (2020), being older works; rather, the claim is that their truth is prima facie in contradiction with computations of the Page curve.

with theorems like (ii) and hence be compatible with GR and prove unitarity at the same time. For the sake of precision, I will focus my analysis on the result of Lesourd (2018), as it involves both a minimal set of assumptions and a minimal notion of causality, which make it a robust result in the field.[3] The reason these two apparently contradictory statements, i.e., the unitarity of black hole evaporation expressed by the Page curve and the failure of black hole evaporation spacetimes to satisfy appropriate causality conditions (which I will introduce later), and hence to be compatible with unitarity, are instead ultimately compatible is that the semiclassical computation of the Page curve involves non-perturbative saddles described by complex, instead of real and Lorentzian, metrics. These are the so-called replica wormholes or replica geometries. Thus, any result in standard Lorentzian geometry cannot apply to such spacetimes. More profoundly, these theorems do not apply to the semiclassical computations of the Page curve because these computations involve a dynamical change in the global structure of spacetime, which is highly problematic (if not straightforwardly impossible) in classical GR. This phenomenon is a new, non-perturbative feature of QG and is crucial to the computation of the Page curve.

As mentioned before, the philosophical interest in black hole evaporation often comes from its ability to illustrate vividly and clarify various foundational issues in the philosophy of physics. Indeed, this chapter is no exception, and

[3] Moreover, compared to most results in the field, Lesourd (2018)'s result, combined with results in Hau et al. (2020), can be extended to AdS spacetimes, thus threatening even the meaningfulness of the original AdS/CFT computations.

the philosophical upshot of its discussion is two-fold. First, it allows me to clearly illustrate an ambiguity in the use of the word semiclassical in both the foundational and high-energy literature and how this ambiguity can lead to confusion and apparent contradictions. My first goal is to point out this ambiguity and explain how to clarify it. In particular, we will see that, in an important sense, both the foundational and the high energy literature are imprecise in their use of semiclassical in this context, though for different reasons. Disambiguating these different uses of semiclassical improves our understanding of the basic structure of the computation of the Page curve. Second, I explain how the discussion of this chapter expands and, in a sense, complements the recent discussion of Weatherall (2022), who identifies various regimes where GR fails, and QG might be required. From this point of view, I highlight that, while Weatherall (2022) only discusses examples of local failures of GR, global failures, i.e., situations where the global structure of spacetime requires QG for its description, are just as significant. Thus, the question of where GR fails and QG is required is, in a sense, much more complex than initially anticipated. The chapter is structured as follows: in §3.1, I introduce the Page curve and its computation via replica wormholes. In §3.2, I introduce Lesourd (2018)'s theorem and explain why it does not threaten replica calculations, relying in particular on Kontsevich and Segal (2021); Witten (2021)'s work on complex metric. In §3.3, I draw some philosophical morals from my discussion of the validity of replica calculations. §3.4 then concludes.

3.1 Replicas and the Page Curve

The goal of this section is to give a brief explanation of the computation of the Page curve from the semiclassical Euclidean path integral. To do so, I start by recalling some facts about the Page curve discussed in the previous chapter (§3.1.1) and then move on to consider the replica trick for computing entropies and, in particular, the role played by replica wormholes in entropy calculations in gravity (§3.1.2). I conclude the section by discussing the Lorentzian analogue of the replica geometry, which will be relevant in what follows (§3.1.3). I will not strive for a complete explanation of these results, as this task would require at least a whole book and would go beyond this chapter's scope. Instead, I will try to explain the basic structure of the computation and refer the reader to the relevant literature whenever possible for a fuller treatment. For a general overview of the subject, the interested reader can consult Almheiri et al. (2021).

3.1.1 The Page Curve

The primary object that we wish to compute here is the Page curve. Its relevance comes from the fact that it describes the change of entropy for a set of unitarily evolving quantum systems. As in the previous chapter, the systems we are interested in are the black hole and its Hawking radiation, i.e., the radiation that the black hole emits as it evaporates.

From chapter 2 we know that to understand the Page curve we can start by taking two quantities:

- The black hole's microcanonical entropy defined as:

$$S_{MC}\left(E(t)\right) = \log \dim \mathcal{H}\left[E(t)\right], \tag{3.1}$$

  where $\mathcal{H}\left[E(t)\right]$ is the Hilbert space of the system at some time t and energy $E(t)$.

- The von Neumann entropy of the radiation defined as:

$$S_{VN} = -\rho_R \log \rho_R \tag{3.2}$$

  where ρ_R is the reduced density matrix of the radiation.

We are interested in the composite quantum state of black hole and radiation. We can assume this state to be pure at the beginning of the evaporation process (to take a somewhat idealized but instructive example, because we start with a shell of matter in a pure state that we collapse to create a black hole). If evaporation is unitary, this state will remain pure; otherwise, evaporation is not unitary. We know by semiclassical considerations that the microcanonical entropy of the black hole, being a function of its area, will decrease as the black hole evaporates. At the same time, since each new quanta of Hawking radiation the black hole emits must be entangled with an interior quanta to preserve smoothness at the horizon, the entropy of the radiation system will increase as the black hole evaporates.

Recall from chapter 2 that, as this process goes on, the two entropies will cross, and at this point, one of two things might happen:

- The entropy of the radiation keeps increasing, even if there are not enough interior modes to purify its state (since the microcanonical entropy is now lower than the radiation entropy). This scenario violates unitarity since it implies that the composite state of black hole and radiation, originally pure, is now mixed.

- The entropy of the radiation starts decreasing, following the microcanonical entropy. This behavior is usually interpreted as the newly emitted radiation being entangled with old radiation, which is consistent with unitarity.

A system of the second kind is said to obey the Page curve (figure 3.1) and is expected to evolve unitarily. The point at which the two curves cross is called the Page time and is expected to be the point at which there is a deviation from Hawking's original computation. Indeed, Hawking's semiclassical computation of black hole evaporation suggests that black holes are systems of the first type and, thus, non-unitary.

This deviation comes from the appearance of new non-perturbative saddles (Almheiri et al., 2020; Penington et al., 2019; Marolf and Maxfield, 2021). The importance of these saddles stems from the fact that their appearance in gravitational path integral computations leads to a modification of the origi-

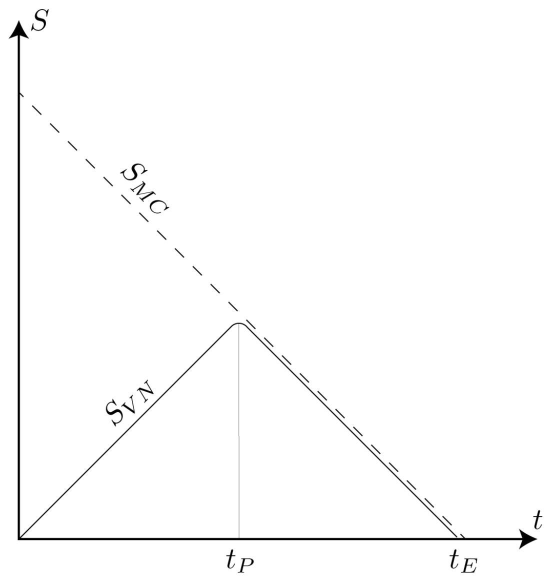

Figure 3.1: The Page curve for an evaporating black hole.

nal entropy calculation, which makes its result compatible with unitarity and allows for the derivation of the Page curve. In the context of holography and AdS/CFT, these new saddles imply that the correct formula for the entropy of the radiation is the so-called island formula (Almheiri et al., 2019; Penington, 2020), which leads to the decreasing entropy expected from systems of the second type.

3.1.2 Replica Wormholes

Let us briefly look at how these new saddles emerge, and thus the role of replica wormholes. First, recall that when we speak of path integral for a gravitational theory, we mean an expression of the form:

$$\langle out|in\rangle = \int \mathcal{D}g\mathcal{D}\phi\ e^{-i\mathcal{S}_{grav}[g,\phi]} \tag{3.3}$$

where $|in\rangle$ and $|out\rangle$ are initial and final states, $\mathcal{S}_{grav}[g,\phi]$ is the gravitational action,[4] and the integral is over the space of metrics g and matter fields ϕ. Now, in general, this expression will be ill-defined, most notably because the measure on the space of fields is not well-defined.[5] As long as we are interested only in a semiclassical approximation, we can still evaluate the path integral using saddle points, as is customary in QFT (Srednicki, 2007). By a saddle point, here we mean a stationary point of the gravitational action, around which we can evaluate the path integral perturbatively. Effectively, the saddle point acts as a vacuum state on top of which we can build perturbations encoding a first approximation to the quantum behavior of gravity.[6] In this way, we can still use our path integral, despite its ultimate ill-definiteness, leading in particular to a powerful semiclassical approximation to the complete QG theory. Let me also introduce a further distinction. By a perturbative saddle (or simply saddle), we will mean a saddle point which is a minimum of the gravitational action. By a non-perturbative sad-

[4] In GR, it would be the Einstein-Hilbert action.

[5] I am here bracketing issues having to do with the renormalizability of gravity, which leads to treating the gravitational path integral as an effective field theory. See Wallace (2021) for a discussion of these issues and an introduction to low energy QG as an effective field theory.

[6] See Wallace (2021) for a discussion of how to derive this saddle point approximation from the path integral.

dle, we will mean instead a saddle which is a stationary point, but not a minimum, of the gravitational action. In particular, amplitudes computed around a non-perturbative saddle are non-perturbatively small, including the amplitude for a transition to a non-perturbative saddle from the perturbative one.[7] The crucial insight behind the semiclassical computation of the Page curve is the inclusion of the non-perturbative saddles in the form of replica wormholes.

The semiclassical computation of the Page curve proceeds through the replica trick, a particularly effective way to compute von Neumann entropies,[8] such as the radiation entropy in which we are interested. The replica trick consists of the following procedure:

- We start by observing that von Neumann entropies are, in general, hard to compute. We would then like to have a more tractable quantity. Such a quantity is the nth Renyi entropy, which we define as follows:

$$\mathcal{S}_n(\rho) = \frac{1}{1-n} \log Tr(\rho^n) \tag{3.4}$$

  for n an integer. We can think of this quantity as computing the entropy between n copies of our original quantum system. In our case, these are n copies of the evaporating black hole and its Hawking radiation. We choose the Renyi entropy because for $n \to 1$, the Renyi entropy coincides with the von Neumann entropy.

[7] An example of non-perturbative saddles are instantons in QFT (Tong, 2005).

[8] See Calabrese and Cardy (2009) for a classic application in $2d$ conformal field theory.

- Having computed the nth Renyi entropy for integer n, we can then analytically continue to real n so that we can take limits of Renyi entropies.

- We can now take the limit for $n \to 1$, which corresponds to the von Neumann entropy. This result is the entropy we were looking for.

Effectively, the replica trick allows us to exchange a difficult computation of the von Neumann entropy with a much simpler computation of the Renyi entropies, from which we can then nonetheless recover the von Neumann entropy we were after.

In the case of a gravitational theory, there are two important points to keep in mind. First, $\mathcal{S}_n(\rho)$ is most easily computed using path integrals in a semiclassical approximation because we can compute traces as

$$Tr(\rho) = \sum_i \langle i|\psi\rangle \langle\psi|i\rangle \tag{3.5}$$

where $|\psi\rangle$ is the initial state of matter which will collapse into a black hole, while $|i\rangle$ is the final state of the radiation after the black hole evaporates (Almheiri et al., 2021). The requirement that the bra and the ket states are the same implements the trace. For the reader familiar with path integrals, this would be an in-in path integral, or timefold (Berges, 2004; Calzetta and Hu, 2008), and $\mathcal{S}_n(\rho)$ comes from applying this notion to multiple copies of the black hole. Thus, the most important step in the replica trick becomes identifying the appropriate saddle point for this computation. Second, when dealing with a theory of gravity, we usually allow saddle points with different topologies, which means that we should allow configurations with possibly very different global structures to contribute to the path integral.

The most obvious saddle point to compute the n-th Renyi entropy is simply a collection of n disconnected black holes (more precisely, n copies of a Schwarzschild spacetime with quantum fields propagating on it, leading to evaporation). We can call this saddle the *Hawking saddle*, since using it inside the replica trick leads to Hawking's original computation of a mixed state for the radiation and non-unitary black hole evaporation. This saddle is not, however, the only possible one. In particular, an equally reasonable alternative to the Hawking saddle is, remembering my previous point on allowing configurations with different topologies into the path integral, a saddle point where instead of the n black holes, we have a shared interior. In other words, a saddle point where wormholes connect the n black holes for which we are computing the Renyi entropy, leading to a connected, instead of disconnected, topology. This difference in topology is the main difference between the connected and disconnected saddles. We call the connected saddle a replica wormhole, and the spacetime geometry of such a configuration a replica geometry (see figure 3.2 for an example).[9] Note that replica geometries will generally be non-perturbative saddle points. The classical description, and thus the perturbative saddle point, for a black hole, is known,

[9]That such spacetimes are stationary points of the gravitational action might not be obvious at first sight. For proof that there is an appropriate gravitational variational problem in which replica geometries are a stationary point, see Colin-Ellerin et al. (2021a).

and it is the Schwarzschild geometry, not a replica geometry. Thus, insofar as replica geometries describe black holes, their contribution is analogous to a non-perturbative effect.[10]

The critical point for our discussion is that by plugging the replica saddle

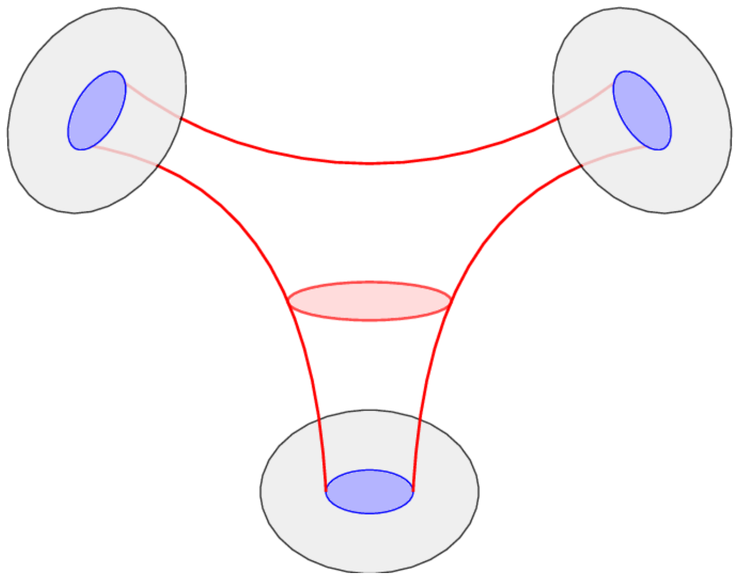

Figure 3.2: A replica saddle for $n = 3$. The three boundary elements are copies of the black hole system, which have a shared interior given by the wormhole.

into the replica trick machinery, the resulting formula for the von Neumann entropy of the radiation follows the Page curve. More precisely, what emerges is that before the Page time, the entropy of the radiation is dominated by the contribution coming from the Hawking saddle, while the replica saddle gives a small contribution. Thus, the entropy is increasing. After the Page time, however, we have a switch in which saddle point dominates the entropy

[10] See Almheiri et al. (2019); Penington (2020) for more on this point.

calculation, and in particular, the replica saddle now gives the largest contribution to the Hawking entropy. This fact means that after the Page time, the entropy of the radiation decreases, which is the behavior predicted by the Page curve. The Page curve is thus derived from a semiclassical approximation within the gravitational path integral. To obtain this result, we have relied heavily on new saddles, characterized by the presence of wormholes, called replica wormholes, connecting the interiors of the various copies of the evaporating black hole.

3.1.3 Lorentzian Replicas

The original replica wormhole computations summarized in the previous section mostly used a gravitational path integral in Euclidean signature.[11] The main reason for this choice is that calculations are much easier in the Euclidean theory compared to the Lorentzian one, as in QFT. Nonetheless, there have been abstract definitions in arbitrary dimensions and multiple concrete constructions, mostly in lower dimensions, of Lorentzian analogues of the Euclidean replica geometries, which give rise to Lorentzian analogues of the original Euclidean computations (Goto et al., 2021; Marolf and Maxfield, 2021; Colin-Ellerin et al., 2021b). It is then helpful to briefly review the structure of replica saddles in arbitrary dimensions and Lorentzian signature, following the one given in Marolf and Maxfield (2021); Colin-Ellerin et al.

[11]The Euclidean results relate to the Lorentzian ones via Wick rotation.

(2021a), since these will be directly relevant to the discussion of theorems on the global structure of black hole evaporation of the next section.

A real-time, i.e., Lorentzian, replica geometry, as defined in Marolf and Maxfield (2021), is represented in figure 3.3 for asymptotically flat spacetimes. Colin-Ellerin et al. (2021a) discuss the same construction for asymptotically AdS spacetimes. For the present discussion, the choice between these two asymptotic behaviors is irrelevant, since the construction of the replica geometry is the same. For ease of exposition, I will work with the asymptotically flat case, since it is more immediately relevant to Lesourd (2018) 's theorem.

To construct a replica geometry, we can start by taking a general relativistic spacetime (M, g)[12] describing an evaporating black hole and selecting a Cauchy surface Σ at a time t after the Page time. We can then identify four regions in this Cauchy surface:

- A region $\partial\Sigma$ corresponding to the intersections between the Cauchy surface and null infinity

- A region Σ_{ext} corresponding, roughly, to the portion of the Cauchy surface sitting outside the black hole horizon

- A region $\mathcal{I}$ corresponding to the portion of the Cauchy surface sitting inside the black hole horizon

[12] Here M is a smooth, Hausdorff n-dimensional manifold, while g is an n-dimensional non-degenerate, real Lorentzian metric. For details on GR and Lorentzian geometry, see Malament (2012); Beem et al. (2017); Carroll (2019).

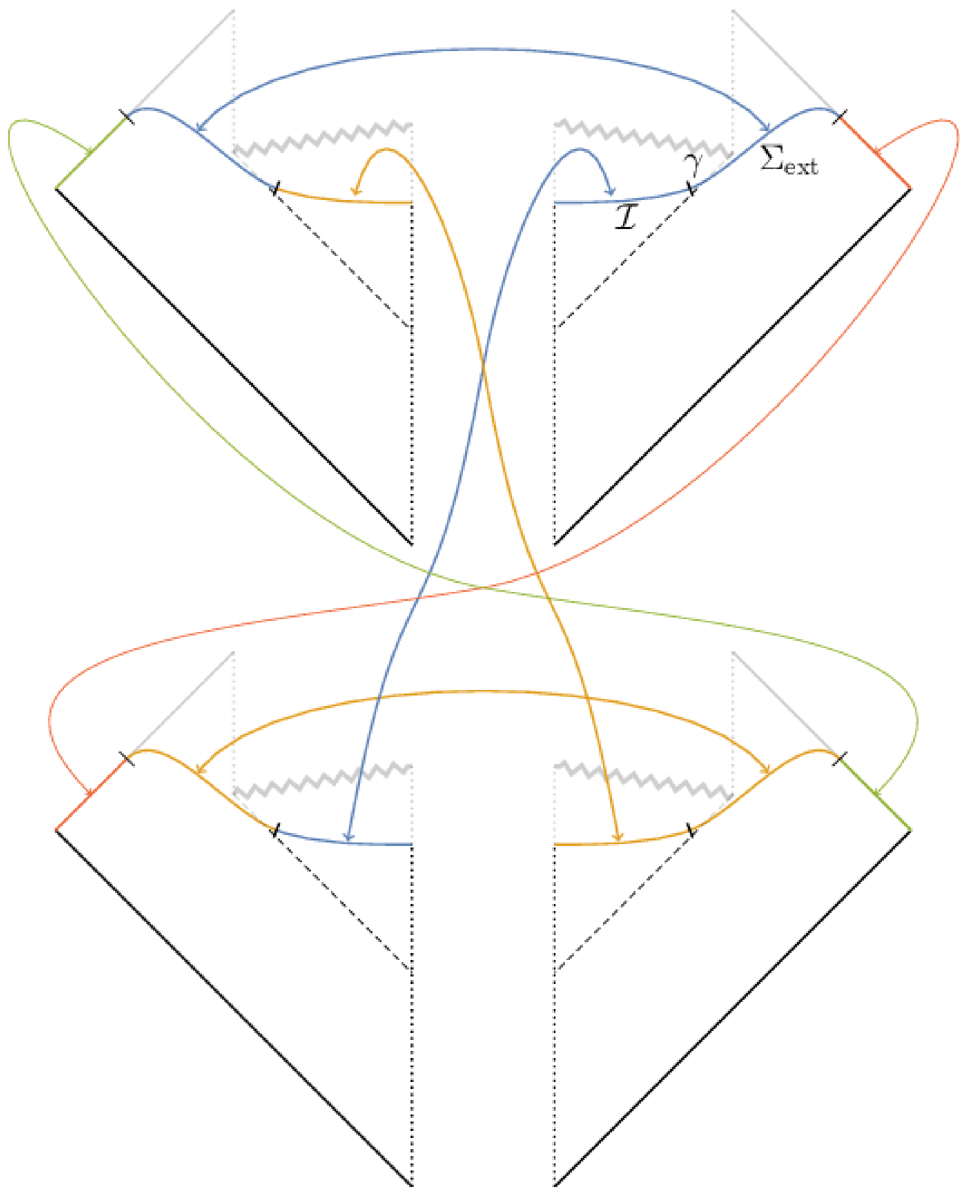

Figure 3.3: An asymptotically flat replica saddle for $n = 2$. The lines show the identifications characteristic of the replica geometry, which are glued together at γ. Two spacetimes at the same height in the picture are bra and ket spacetimes, for which Σ_{ext} is identified (blue and yellow horizontal arrows). Spacetimes at different heights are part of different replicas, for which $\mathcal{I}$ is identified (blue and yellow vertical lines). The green lines give the analogous identifications for the boundary $\partial\Sigma$.

- A codimension two region γ, which corresponds to the boundary between $\mathcal{I}$ and Σ_{ext}

Given these four regions, we can now proceed to construct a replica geometry. To start, take $2n$ copies of our original spacetime (M, g). $2n$ copies, instead of n, since, as noted in the previous section, to compute the Renyi entropies, we need to perform a timefold or in-in path integral (Berges, 2004; Calzetta and Hu, 2008). Timefolds are required whenever we wish to compute gravitational entropies and quantities in real-time thermal physics. A timefold is an expression of the following sort:

$$\langle\psi| e^{iHt} A(t) e^{-iHt} |\psi\rangle \tag{3.6}$$

for ψ a quantum state and A an operator.[13]

From a path integral perspective, we can interpret (3.6) as first propagating ψ forwards in time, then inserting A, and then propagating backwards in time. From the spacetime viewpoint, timefolds involve two copies of spacetime, called *bra* and *ket* spacetimes, one for $\langle\psi|$ and one for $|\psi\rangle$. In the replica geometry, these spacetimes are identified across Σ_{ext} to ensure that the bra and the ket states are the same, i.e., that they are both $|i\rangle$ states of the radiation in (3.5).[14] We need to glue these spacetimes together to get

[13] In the present context A would be an operator suitably related to the entropy of the system, such as the *swap operators* of Marolf and Maxfield (2020, 2021).

[14] One also requires that time flows in opposite directions in the bra and ket spacetimes, leading to further complications in the spacetime structure at γ. For ease of exposition, I will mostly ignore these complications in the following sections, as my arguments do not

a replica geometry. Recall that a replica geometry is one where we identify the interiors of the $2n$ black holes. Calling each spacetime involved $M_i^{b/k}$,[15] where b and k stand for whether it is a bra or ket spacetime respectively, and $i = 1, \ldots, n$ stands for which replica it is, we identify Σ_{ext} in M_i^k with Σ_{ext} in M_i^b, and $\mathcal{I}$ in M_i^k with $\mathcal{I}$ in M_{i-1}^b. We are gluing these spacetimes together at γ (see figure 3.3). The same identifications should be extended to each replica's boundary ∂M. This geometry, let us call it $\hat{M}$, is called, following Colin-Ellerin et al. (2021a), a Lorentzian replica wormhole.

Colin-Ellerin et al. (2021a) show that the resulting spacetime has some remarkable properties:

- It has a $\mathbb{Z}_{\ltimes}$ symmetry, called replica symmetry, which means that we can exchange the ith replica with the $i-1$th.

- It has a CPT symmetry, implementing the exchange between the bra and ket spacetimes in the timefold.

- Most importantly for our purposes, it is highly singular at γ (Marolf and Maxfield, 2021). This singularity is due to γ having multiple past lightcones, one for each of the $2n$ spacetimes involved in this construction.

This last property is the one most relevant to my discussion. Its origin is

hinge on this fact for two reasons. First, timefolds seem to be mostly calculation devices to compute entropies in this context, and thus their physical significance is unclear. Moreover, as shown in Witten (2021), timefolds satisfy the allowability criterion discussed in the next section, and thus their presence does not change my main conclusions.

[15] I suppress mention of the metric g from the notation to avoid clutter.

due to the peculiar topology of the replica geometry compared to the Hawking saddle. Where the latter has a disconnected topology, the former has a connected one. Connectedness leads to the appearance of the singularity at γ since it implies that the copies of the black hole are glued at γ, leading to γ having multiple past lightcones. To see the importance of the singular behavior at γ, note that when computing the path integral, we need to integrate across γ, for otherwise, we would not be able to integrate across different replicas. Intuitively, the singularity at γ would stop the integration contour from going from one replica to the other. We thus need to regularize spacetime at γ, and, as observed by Marolf and Maxfield (2021); Colin-Ellerin et al. (2021a), we can do so by complexifying the metric at γ.[16] In this way, we can have a smooth spacetime that we can integrate over at γ, at the price of renouncing to only deal with real Lorentzian metrics. Let me briefly note, however, that the appearance of complex metrics as saddles is not immediate grounds for dismissing these constructions as unphysical. Indeed, complex metrics often appear in perfectly respectable path integral computations, as for example in the semiclassical path integral for the Kerr black hole of Gibbons and Hawking (1993) (see also Gibbons et al. (1978);

[16]By complex metric I mean a symmetric $\mathbb{R}$-bilinear form (Kontsevich and Segal, 2021). The allowability criterion discussed in the following section can be seen as an extension to complex metrics of the usual condition of positive-definiteness for the metric (Witten, 2021). An example of complexification is the use of an $i\epsilon$ prescription (Kontsevich and Segal, 2021; Witten, 2021). The application of an $i\epsilon$ prescription to spacetime is the move from the n-dimensional Lorentzian metric $ds^2 = -dt^2 + \sum_{i=2}^{n} dx^2$ to the complex metric $ds^2 = -(1 \pm i\epsilon)dt^2 + \sum_{i=2}^{n} dx^2$. For more on complex metrics and their properties in this context see Kontsevich and Segal (2021); Witten (2021).

Halliwell and Hartle (1990); Louko and Sorkin (1997) for further examples). Complex metrics are quite different from standard real Lorentzian metrics: for example, they can be used to construct well-behaved topology-changing spacetimes (Louko and Sorkin, 1997), which is quite hard to do with real metrics.[17]

This discussion's upshot is that the replica geometries $\hat{M}$ are complex metrics, not real ones. Indeed, the appearance of these complex metrics will allow replica computations to avoid Lesourd (2018) 's theorem, as we will see in the following section.

3.2 Causality Conditions and Complex Metrics

In this section, I will give a brief introduction to Lesourd (2018) 's theorem and explain why it is problematic for replica computations (§3.2.1). I will then explain why its conclusion ultimately does not apply to these computations, which thus preserve their generality (§3.2.2).

3.2.1 A Theorem on Causal Continuity

Lesourd (2018) 's theorem can be expressed through the following statement:[18]

[17]See Earman (2008) for a review of the issues regarding topology change in GR.

[18]Note that from now on spacetimes (M, g) will be four-dimensional, and not n-dimensional as in the previous section.

Theorem. *Let (M, g) be a spacetime asymptotically flat at null infinity, such that its conformal boundary consists of two disconnected null components $\mathcal{J} \equiv \mathcal{J}^+ \cup \mathcal{J}^-$ each having topology $V \times \mathbb{R}$*[19]*. Suppose that the following properties obtain:*

- *there is a non-empty event horizon $\partial I^-(\mathcal{J}^+)$*[20] *and a non-empty black hole region defined by $B \equiv I^+(\partial I^-(\mathcal{J}^+))$ such that $\partial B = \partial I^-(\mathcal{J}^+)$ and $B \cap I^-(\mathcal{J}^+) = \varnothing$,*
- *$\partial I^-(\mathcal{J}^+) \subset I^-(\Sigma)$ where Σ is a complete cross-section of $\mathcal{J}^+$, i.e., a spacelike embedded submanifold of $\mathcal{J}^+$ with topology V.*

Then (M, g) is causally discontinuous.

The first condition serves to establish that we are dealing with a black hole spacetime, while the second condition establishes that the black hole fully evaporates, since it means that there is an observer at null infinity for whom the horizon is in their causal past. The relevant causality condition is causal continuity. Let us define it. First, we need the notion of spacetime being reflecting, defined thus:

- A spacetime (M, g) is said to be past reflecting if $I^+(p) \subset I^+(q) \rightarrow I^-(q) \subset I^-(p)$. Future reflecting is defined dually. A spacetime is reflecting if it is both future and past reflecting.

We also need the notion of spacetime being distinguishing:

[19] We can think of V as a spatial slice of $\mathcal{J}$.

[20] $I^{-(+)}(p)$ is the chronological past (future) of $p \in M$.

- A spacetime (M, g) is distinguishing iff, for all $p, q \in M$, $I^{-}(p) = I^{-}(q)$ or $I^{+}(p) = I^{+}(q)$ implies $p = q$.

With these notions in hand, we can then define the notion of causal continuity:

- A spacetime (M, g) is causally continuous iff it is distinguishing and reflecting.

Intuitively, causal continuity encodes the idea that, given an observer at p, we should not be able, by slightly displacing them from p, to alter the contents of their past lightcone drastically.

Remember that we can extend this result to AdS spacetimes by combining it with the results of Hau et al. (2020). In the AdS case, the relevant causality condition is called globally-hyperbolic-with-boundary. In what follows, I will stick with the asymptotically flat case.

The interest in this theorem comes from two points: (i) causal continuity is a weaker condition than global hyperbolicity, which is the causality condition usually shown to be incompatible in theorems on causality and black hole evaporation (Kodama, 1979; Wald, 1984; Manchak and Weatherall, 2018); (ii) the theorem does not make any assumption regarding the structure of spacetime at the singularity. The only assumptions concern the presence of the black hole and its evaporation. Together, these two claims lead to a more general result than its predecessors.

As usual with theorems of this kind, Lesourd (2018) 's theorem aims to show that the causal structure of evaporating black holes is too pathological to allow for a well-defined and predictable notion of evolution for quantum fields. Insofar as this is true, it would be incompatible with claims that black hole evaporation is unitary, since unitarity implies a well-defined and predictable notion of evolution for quantum fields in the black hole spacetime. Let me briefly expand on this point. First, we need to understand how to apply this theorem, stated in the abstract language of Lorentzian Geometry, to the physical description of an evaporating black hole, which involves quantum fields and a semiclassically quantized gravitational field. The tool to bridge this gap is the semiclassical Einstein field equation (SEFE), which reads thus:

$$G_{\mu\nu} = \frac{8\pi G_N}{c^4} \langle T_{\mu\nu} \rangle \tag{3.7}$$

where $G_{\mu\nu}$ is the Einstein tensor, $\langle T_{\mu\nu} \rangle$ the expectation value of the stress-energy tensor, and G_N Newton's constant.

The SEFE can be derived from the gravitational path integral, roughly by taking the metric g giving rise to a specific $G_{\mu\nu}$ as the perturbative background.[21] The main advantage of this equation compared to the standard Einstein field equation is that on the right-hand side, we have the expectation value of the stress-energy tensor $T_{\mu\nu}$, which means that the field content of the spacetime is given by quantum, instead of classical, fields. At the same time, g is still treated as a classical field, which means that notions

[21]See Wallace (2021) for more on the relation between SEFE and the gravitational path integral.

such as causality or energy conditions, and theorems such as Lesourd (2018) 's, should apply. Thus, using the SEFE, we can represent gravitational situations where the quantum nature of the fields involved is essential, as in black hole evaporation. Moreover, we can export the precise notions and results obtained on the geometry of general relativistic spacetimes to these situations.

With this in mind, a helpful example to understand why the failure of causal continuity is problematic is given by Anderson and DeWitt (1986) (see also Dowker and Surya (1998) for further discussion). There, they show that a particular spacetime violating causal continuity[22] develops, at the point where causal continuity fails, a burst of infinite energy, which makes any reasonable notion of evolution impossible. In particular, this burst of infinite energy implies the impossibility of defining a path integral describing the evolution of quantum fields.[23] This kind of worry seems to be the central point of Lesourd (2018) 's theorem, i.e., that failure of causal continuity means that there is no predictable, well-defined evolution for quantum fields. In path integral language, without causal continuity, there is no well-defined path integral.

[22] The trouser spacetime describing topology change from two disconnected circles to a single one.

[23] These kinds of examples later led Sorkin to conjecture that only causally continuous topology-changing spacetimes admit well-defined quantum evolution and thus path integrals (Dowker and Surya, 1998; Borde et al., 1999).

3.2.2 Complexity, Acceptability, and Replicas

Having seen why theorems on the causal structure of spacetime might apply to replica wormhole calculations, it is now time to see why they probably do not.

First, note that it is not sufficient to say that in the context of replica wormhole calculation, we are dealing with an exchange between two different saddles in the gravitational path integral; this fact by itself is not sufficient to resolve our conundrum. Straightforwardly, this fact implies that there is no description of the spacetime involved via a single metric and, thus, no complete description of the replica calculation in terms of the SEFE. Consequently, Lesourd (2018) 's theorem would appear ill-suited to apply to this situation. Indeed, the exchange in saddles is ultimately why replica calculations differ from the semiclassical scenarios to which such theorems apply. The reason is that the topology change involved in moving from the Hawking saddle (disconnected topology, semiclassical result) to the replica saddle (connected topology, Page curve) leads to the appearance of complex metrics which, as we will now see, avoid the pitfalls of Lesourd (2018) 's theorem. Thus, ultimately, it is the change in the global structure of spacetime, in particular in its topology, generated by the exchange between the Hawking and the replica saddles, that helps avoid the contrast between this computation and the theorems on the causal structure of evaporating black holes. Nonetheless, simply pointing to this exchange in saddles is not enough, and some more careful argument is needed. Indeed, in principle, one still runs the risk that the theorem applies to spacetime $\hat{M}$, i.e., to the replica spacetime valid after the Page time. We would then find ourselves with a sort of *post-Page-time causality trouble*, where we restrict ourselves to the spacetime after the Page time and find that the violation of causality relevant to black hole evaporation would still take place,[24] getting us back to square one, where the generality of the calculation is in danger. Therefore, we need to find a reason why $\hat{M}$ specifically avoids the unpalatable consequences of Lesourd (2018) 's theorem.

To do so, start from the observation that Lesourd (2018)'s theorem as stated only applies to *real* Lorentz metrics, as is evident from the definitions in (Lesourd, 2018, pp. 4-5) and especially in (Beem et al., 2017, pp. 20-25), which gives the basic geometric background to Lesourd (2018). However, replica geometries are not real Lorentz metrics but complex metrics! Thus, quite straightforwardly, the theorem, as stated, cannot apply to them. As I noted above, the appearance of complex metrics, and thus the avoidance of Lesourd (2018) result, is ultimately the product of the change between the Hawking and replica saddles. However, it is the complexity of the metric that is crucial to this specific argument. Indeed, as I mentioned in the previous section, complex metrics have quite significant differences from real Lorentz metrics, so we should not expect notions and results developed for the real context to extend seamlessly to the complex case, at least not without ex-

[24]Though it would not concern the entire black hole evaporation spacetime as in the original statement of the theorem.

plicit proof. Lacking such explicit proof, we have no reason to doubt the generality and applicability to dimensions beyond two, and four dimensions in particular, of the replica calculations of the Page curve, and thus on the surprising and deep consequences of these results.

Can we do more than this and show that results analogous to Lesourd (2018) cannot be proven for replica geometries? Such arguments would require an explicit description of the geometry of a four-dimensional replica spacetime, which, to my knowledge, is still lacking.[25] However, we can at least gather some evidence in favor of the claim that replica geometries do not pose a threat to well-defined quantum evolution by looking at analogues of replica wormholes, along the lines of the approach used in Wald (1984) to discuss the predictability of quantum evolution at the end of black hole evaporation. Note that well-defined quantum evolution should be enough to ensure that even if an analogue of Lesourd (2018)'s theorem exists for replica geometries, it is not a threat to the replica calculations of the Page curve. Indeed, whether well-defined quantum evolution implies causal continuity is, to some extent, besides the point. The crucial issue is the well-definiteness of quantum evolution and, thus, unitarity, not causal continuity per se. Rather, causal continuity is supposed to be a condition for well-defined evolution. Hence, proof of well-defined quantum evolution entails that either replica geometries are causally continuous or that, contra Lesourd (2018), causal continuity is not required for well-defined evolution. Further work will be required to

[25] Though see Colin-Ellerin et al. (2021b) for recent progress on this issue.

prove beyond doubt that replica geometries allow for well-defined evolution, and to decide between these two options. For this chapter, I will limit myself to presenting evidence favoring the hypothesis that replica geometries allow for well-defined quantum evolution.

First, we need a criterion for the well-definiteness of quantum evolution for complex metrics, since this is ultimately at stake in Lesourd (2018)'s theorem and is most relevant for replica calculations. A promising criterion has recently been proposed by Kontsevich and Segal (2021), and further expanded upon by Witten (2021).[26] The criterion effectively amounts to a positivity condition on complex metrics.More formally, if we represent an n-dimensional complex metric g as a matrix, its n complex eigenvalues λ_i satisfy (Kontsevich and Segal, 2021, Thm. 2.2):

$$\sum_{i=1}^{n} |\arg \lambda_i| < \pi \tag{3.8}$$

A complex metric satisfying this condition is an allowable complex metric. Real Lorentzian metrics live on the *boundary* of the space of allowable complex metrics.

The claim of Kontsevich and Segal (2021) is that this condition should be taken as an axiom for QFT and effectively tell us when a QFT is well-defined in a given spacetime. Witten (2021) extends this claim to semiclassical gravity via the requirement that an admissible saddle point for the gravitational

[26] See also Visser (2022) for further discussion.

path integral should allow for a well-defined QFT propagating on it, where well-defined QFT means satisfying (3.8). Why should (3.8) be enough? Because satisfaction of (3.8) implies that the path integral for an arbitrary p-form field[27] converges in the complex spacetime under study (Kontsevich and Segal, 2021). Thus, for a large and significant[28] class of quantum field theories, satisfaction of (3.8) amounts to the existence of a perfectly well-defined and predictable notion of evolution. Indeed, this evolution is nothing more than the standard path integral evolution that we are familiar with from QFT, outputting transition amplitudes and expectation values conforming to the rules of quantum mechanics as results.[29]

It is then immediate to see why (3.8) should be relevant to our purposes, for satisfying this condition would mean that replica geometries admit a standard and predictable notion of evolution for (a large class of) quantum fields. While, as mentioned at the beginning of the section, definitive proof of the satisfaction of (3.8) for replica geometries is not currently available[30], we can look at similar geometries and check it for them. In particular, a useful example is provided by the Lorentzian double cone geometry (Saad et al., 2018), which is a type of Lorentzian replica wormhole for n=2 replicas, and whose relation to (3.8) has been studied in Witten (2021). While this geometry

[27]A p-form field is a field which is locally a p-form.

[28]Significant because many of the known well-defined quantum field theories are of this form (Witten, 2021).

[29]For more on this criterion as an axiom in QFT see Kontsevich and Segal (2021); Witten (2021).

[30]Since a four-dimensional asymptotically flat explicit definition of a replica geometry is not available.

has not had a significant role in the computation of the Page curve, it has played an essential role in the computation of a different quantity relevant to quantum black holes, the spectral form factor. This quantity is related to the black hole's late-time behavior and, in particular, to the manifestation of chaos in this regime. We define the Lorentzian double cone as follows:

$$ds^2 = -\sinh^2 r \; dt^2 + dr^2 \tag{3.9}$$

where t is a real-time coordinate satisfying $t \cong t + T$, while r is a real spatial coordinate.

This metric is singular at $r = 0$. We can regularize it through complexification, and in particular, by applying the following condition:

$$r = u \; - \; i\epsilon, \; u \in \mathbb{R} \tag{3.10}$$

The resulting metric is:

$$ds^2 = -\sinh^2 (u \; - \; i\epsilon) dt^2 + du^2. \tag{3.11}$$

As shown in Witten (2021), this metric does indeed satisfy (3.8). This result provides the first evidence that replica geometries allow for the causal propagation of quantum fields. Indeed, as the double cone is one of the most basic replica geometries, and various replica geometries are just generalizations of the double cone, this argument strongly suggests that replica geometries allow for predictable path integral evolution.

A slightly more general argument goes as follows. Colin-Ellerin et al. (2021a) point out that the metric of a replica geometry at γ, i.e., where the metric is complexified to remove the singularity from replica wormholes, is analogous to a class of metrics studied in Louko and Sorkin (1997). In particular, they apply the same methods of Louko and Sorkin (1997) to regularize/complexify these metrics to compute their contribution to the gravitational action, which is given by a (generalization of) the Gauss-Bonnet theorem.[31] As shown in Witten (2021), the metrics discussed by Louko and Sorkin (1997) satisfy (3.8). Indeed, as Witten (2021) points out, the Gauss-Bonnet theorem gives its standard value for allowable complex metrics but not general complex metrics. Both the metric around γ (Colin-Ellerin et al., 2021a) and the metrics studied by Louko and Sorkin (1997) obey the Gauss-Bonnet theorem, thus lending further support to the claim that replica geometries are sufficiently casually well-behaved to allow for path integrals.

While, of course, none of these considerations is by itself proof of replica geometries being immune from theorems such as Lesourd (2018)'s, they provide good evidence that any causal pathology in replica geometries should not threaten the possibility of well-defined quantum evolution as given by

[31] The Gauss-Bonnet theorem states that, given a two-dimensional manifold M with boundary ∂M, and for K the Gaussian curvature of M and k_g the geodesic curvature of ∂M, $\int_M K\, dA + \int_{\partial M} k_g ds = 2\pi\chi(M)$. Here dA is the area element of M, and ds the line element along ∂M. $\chi(M)$ is the Euler characteristic of M, a topological invariant. Thus, the Gauss-Bonnet theorem relates curvature with topology. In replica geometries, the Gauss-Bonnet theorem applies since the metric around γ is effectively two-dimensional, since curvature only propagates along two dimensions there (Colin-Ellerin et al., 2021a).

path integrals, i.e., the sense relevant to these calculations. Unless a specific theorem relevant to the replica computations is proven, the burden of proof lies on those skeptical of their generality.

3.3 Replica Philosophy

Having discussed why theorems on causal structure do not straightforwardly apply to replica calculations, let me now come to the philosophical payoff of this discussion. Indeed, while the ultimate fate of causal propagation in evaporating black holes is still an open question, the discussion thus far allows me to highlight various interesting conceptual and foundational issues. In particular, I will focus on the meaning of semiclassical limit in light of this discussion (§3.3.1) and on the regimes where GR fails (§3.3.2).

3.3.1 What Is Semiclassical?

Let us start with a straightforward puzzle that seems to emerge from the previous discussion. Both Lesourd (2018) 's theorem and the replica computations claim to be semiclassical results, but at the same time, they seem to imply very different results regarding the behavior of black holes during evaporation. We have seen that the origin of this discrepancy lies in the replica computations taking into account new non-perturbative complex saddles that become dominant in the path integral after the Page time. However, if both results are valid in the same regime, i.e., semiclassical gravity, then why did Lesourd (2018) 's theorem not consider these saddles? More generally, what do those working in high energy physics and those working on the foundations of GR mean by semiclassical that leads them to such different results? This question is crucial if we think semiclassical gravity results proving specific hypotheses about QG (such as the unitarity of black hole evaporation) should give greater evidential support to those hypotheses. Without knowing what semiclassical gravity means, we cannot decide whether it should give us greater confidence in specific hypotheses.

Prima facie, it seems that the two approaches should agree. Lesourd (2018)'s theorem functions as a semiclassical result insofar as we are using the SEFE, as explained in the previous section, to define our semiclassical approximation. On the other hand, the replica computation proceeds by taking the gravitational path integral[32] and applying a saddle point approximation. This approach is standard for defining a semiclassical approximation when using path integrals, since it allows for a perturbative evaluation of the path integral around a stationary point of the action. However, this saddle point approximation leads to a description equivalent to the SEFE, as long as we ensure that the metric g appearing in the SEFE is the same as the saddle point metric. The SEFE description then emerges as the leading term in the perturbative expansion around the saddle point.[33] Therefore, prima facie, the two approaches should agree; however, they do not.

[32]Treated as an effective field theory.

[33]Indeed, as mentioned before, this is the way to recover the SEFE from the path integral description.

The difference between the two results lies in the path integral expression including, in principle, all sorts of non-perturbative effects that the SEFE description, effectively the first term in a perturbative expansion, cannot see. One of these non-perturbative effects is the exchange between the standard saddle and the replica saddle that takes place around the Page time. To see why, note that the exchange of saddles cannot be studied in a perturbative expansion around any saddle since it involves the change of the relevant saddle. Hence, it must be a non-perturbative effect. This exchange leads to the difference between the two approaches, since the replica saddle with its complex geometry leads to the Page curve. Moreover, the replica geometry itself only makes sense as a non-perturbative effect since, while it is a stationary point of the gravitational action (Colin-Ellerin et al., 2021a), it is not a minimum of the action, which in the case of a (non-rotating, not charged) black hole is the Schwarzschild geometry. Indeed, as discussed in Almheiri et al. (2020); Penington et al. (2019), the contribution of the replica geometry to the gravitational path integral is usually subleading and comes to dominate after the Page time only because of non-perturbative corrections.

The observation that the difference between the result of Lesourd (2018) and the replica computation lies in the ability of the path integral to take into account non-perturbative effects suggests a necessary clarification regarding the meaning of semiclassical in this context. When employing the SEFE to describe QG's semiclassical approximation, one must be extremely careful regarding the possibility of non-perturbative effects becoming relevant at some point. Moreover, note that this in and of itself does not imply a failure of the semiclassical approximation: the replica computation is semiclassical since it is carried out in perturbation theory around a saddle point, without being treatable entirely in terms of the SEFE, since it does not allow us to take into account the exchange among saddles. This fact is particularly relevant for the foundational literature on GR and semiclassical gravity, where it is common to rely on the SEFE and related results in Lorentzian geometry to derive precise results regarding the behavior of semiclassical gravitational phenomena such as black hole evaporation. While in principle correct, this approach risks being physically irrelevant if care is not taken to ensure that energies are low enough to allow for a semiclassical treatment and that a perturbative approximation around a single saddle point is possible. Otherwise, a description fully in terms of the SEFE is not valid, as happens in the case of black hole evaporation from the perspective of the replica calculations. Conversely, when relying on path integrals to derive results about semiclassical physics, one should be just as careful regarding the extent to which the results themselves do indeed count as semiclassical. As we have seen, effectively within a semiclassical approximation, relying on perturbation theory around saddle points, one can derive highly non-trivial results involving considerations from non-perturbative physics in the form, for example, of saddle changes and non-perturbative saddles. In the context of black hole evaporation, at least, the label semiclassical is appropriate since the exchange of saddles is because the leading term in the semiclassical approximation to the path integral changes. Thus, the exchange of saddles keeps track of changes in the semiclassical description. Nonetheless, the procedure is ultimately non-perturbative, and there is no way to describe the change from the Hawking saddle to the replica saddle in semiclassical terms. Instead, the change of saddles has to be assumed and possibly justified from semiclassical considerations, but not derived, being a non-perturbative effect. However, while in black hole evaporation use of these methods might not break the validity of the semiclassical approximation, it is not guaranteed to not do so in other contexts, and much care should be exercised to avoid smuggling in assumptions about full QG into semiclassical computations. Indeed, this care is crucial if these computations are supposed to gain some hypothesis about QG a greater degree of support by relying on it being verified already at the semiclassical level.

A valuable way to keep track of these distinctions might be Wallace (2021) 's distinction between low energy QG and semiclassical gravity, where low energy QG involves gravitational path integral computations of low energy amplitudes in quantum GR treated as an effective field theory. In contrast, semiclassical gravity means computations of gravitational and QFT quantities via the SEFE. As this distinction can neatly separate between these two theoretical contexts, counting the replica computations as low-energy QG computations might be conceptually helpful, at least insofar as it avoids possible confusions between the specific regime in which a certain computation takes place, and thus confusions regarding which results are relevant to the computation under study. Given, however, the large degree of overlap between these two contexts, care must still be taken regarding their interaction, and in particular, regarding the evidential value of computations in low energy QG, which might coincide with that of a semiclassical gravity computation,[34] or might be lower, depending on the specifics of each computation.

3.3.2 On Regimes Where General Relativity Might Fail

Let me briefly comment on the replica computations' relevance for the question of when QG effects become relevant, or equivalently when GR fails. This question was considered in detail in Weatherall (2022), where two points are stressed. First, no quantity precisely captures the expectation that GR will fail at high energies. The closest one is using curvature scalars to capture the high energy regime where GR fails. Second, Weatherall (2022) gives a description of two important examples of physical situations where GR might fail and a discussion of whether they involve unbounded curvature scalars. These examples are singularities and violations of the strong cosmic censorship conjecture.

From the preceding discussion, it is immediate to see that replica computations provide an interesting extension of both points discussed by Weatherall (2022). Regarding the second point, replica wormholes provide an interesting example of a regime where classical GR fails, and quantum effects must be

[34] As seems to be the case for replica computations from the previous discussion.

considered. More interesting is the relation of replica geometries to Weatherall (2022) 's first point, i.e., the use of curvature scalars to define where GR should fail. Indeed, it seems that for replica computations, there is no sense in which the failure of GR is realized by a curvature scalar blowing up. As explained in §3.1.3, replica geometries deviate from standard Schwarzschild geometries at γ, i.e., at the horizon, where the various replicas are glued together. The horizon of a black hole is a low curvature region of spacetime, and, owing to the equivalence principle, nothing dramatic happens there. Nonetheless, it must be the case that something strange does happen in the end, and a deviation from GR, leading to the Page curve, takes place.

As mentioned in §3.1, what ultimately drives the deviation from GR in replica computations is the difference in the global structure, and in particular topology, between the Hawking and the replica saddle. These processes are generally difficult to describe in GR, while they seem to play an important role in QG, even beyond the replica calculations.[35] The crucial issue for the present discussion is that if topology is what makes a difference and leads to the failure of classical GR in replica calculations, then topology cannot be captured via a quantity such as a curvature scalar. The reason is that topology is a global quantity, which cannot be fixed solely in terms of local quantities such as curvature scalars.[36] Thus, we seem to have found not sim-

[35]See DeWitt and DeWitt (1964); Hawking (1978); Iqbal et al. (2008); Penington and Witten (2023) for concrete examples.

[36]For local structure, I follow Manchak (2009, 2013)'s definition, where local structure is defined as comprising any property that, given two spacetimes (M, g) and (M', g'), is preserved whenever they are locally isometric. Two spacetimes are locally isometric iff,

ply a class of examples where GR fails, but a class of examples where GR fails in a significantly different way from those contemplated in Weatherall (2022). In particular, the failure of GR involved in the replica calculation has to do with the global structure of GR rather than with any property that can be defined locally. This fact signals one way QG is expected to differ most radically compared to classical GR: while in GR, global properties are, for the most part, unchanging features of spacetime, they can be dynamical in QG. In particular, it is expected that QG might display features such as topology change (Brennan et al., 2017; Maldacena et al., 2021; Penington and Witten, 2023), or even changes in the signature of spacetime (Hartle and Hawking, 1983; Dijkgraaf et al., 2018; Brahma, 2017). If this were the case, then the scale of situations where failures of GR might pop up would be much greater, since changes in the global structure of spacetime are not fully bound by local quantities such as energy or curvature and thus cannot be relegated to regimes where those quantities blow up. In other words, we could have failures of GR even in regimes where GR seems to be valid, at least from the perspective of an observer who only has operational access to local spacetime structure, since the failures do not concern the local but rather the global structure of spacetime. Indeed, black hole evaporation and the replica computation of the Page curve offer a particularly striking exam-

for all $p \in M$, $(O, g_{|O})$, for O a neighborhood of p, is isometric to an open set $(O', g_{|O'})$ in M', and vice versa when exchanging (M, g) and (M', g'). A global property is not fixed for locally isometric spacetimes. In this sense, curvature scalars are part of the local structure of spacetime, while topology is not. For an alternative definition of local and global structure in GR, see Geroch (1970); Krasnikov (2014).

ple of this phenomenon.

3.4 Conclusions

We have seen in this chapter how replica wormhole computations of the Page curve are compatible with, and avoid the consequences of, theorems on the causal structure of evaporating black holes, focusing in particular on a result of Lesourd (2018). Complex saddles in the gravitational path integral, which in turn are a consequence of the changes in the global structure of spacetime due to QG, play a crucial role. Since results in the causal structure of spacetime are stated for real Lorentz metrics, they do not apply to replica wormholes. Indeed, spacetimes analogous to replica wormholes show a well-behaved causal structure that allows for standard path integral evolution.

This discussion raises two interesting philosophical points: first, there is some ambiguity in the physics and philosophy literature regarding the use of the term semiclassical when dealing with computations regarding evaporating black holes, which is relevant to the status of hypotheses about QG and in particular to the relevance of various GR theorems for said hypotheses. I have suggested that much care should be taken in resolving these ambiguities on a case-by-case basis. Second, replica computations provide an interesting new example of a deviation from classical GR due to QG. In particular, this deviation is due not to any local property of spacetime, but rather to its global properties and QG's influence on them.

Concluding in this way the discussion of these first two chapters of the foundations of AdS/CFT models of quantum black holes and their path integral derivation, we are now ready to move to the more metaphysical discussion of the following chapters. We will start in chapter 4 with a discussion of how Humeanism can be adapted to QG in the context of these models.

Chapter 4

Amplitudes And Humeanism In Quantum Gravity

After the more foundational discussion of the previous two chapters, I start in this chapter a more metaphysics based of the AdS/CFT models of quantum black holes. To do so, I begin by looking at the issue of Humeanism in QG.

Of all grand theses in metaphysics, Humeanism is perhaps one of those which have had the most significant interaction with modern theories of physics. This interaction owes in particular to the difficulties faced by Humeanism in the face of quantum entanglement (Maudlin, 2007), which seems to threaten its assumption of **separability**. More recently, a possibly even more troubling issue for Humeans has emerged from discussions on QG foundations. This issue stems from the alleged disappearance of spacetime that is common to many approaches to QG (Lam and Wüthrich, 2021a). In particular, insofar as Humeans ordinarily assume that laws of nature, and more generally all modal matters, can be analyzed in terms of local matters of fact organized in terms of spatiotemporal relations, the non-existence of spacetime itself seems to pose a grave threat to the meaningfulness of the view itself. Urgent work in rectifying this state of affairs is needed, for otherwise, Humeanism would be doomed on purely scientific grounds.

This chapter aims to articulate a version of Humeanism that is compatible with QG and thus does away with spacetime relations in articulating the supervenience basis. The relation that I will propose to substitute spacetime and thus do this job is that of *having non-zero scattering amplitude*, a relation that has a natural definition in most QG approaches and that nonetheless has two attractive features:

- It does not necessarily rely on spacetime for its definition but is nonetheless a natural external relation, thus suitable to structure the Humean supervenience basis.

- It is sufficient to recover, in the appropriate limit, the structure of classical spacetime with quantum fields propagating on top of it, i.e., semiclassical gravity. Thus, it allows us to apply the standard Humean story in the limit where spacetime reappears and account for all the other less fundamental facts in this way.[1]

[1]Modulo issues with quantum entanglement, though see Cinti and Sanchioni (2021b) for how semiclassical gravity, coupled with the ER=EPR conjecture, might help with these

The first feature will be discussed in general in this chapter, while the second will be discussed in a particular example derived from AdS/CFT.

While inevitably reliance on AdS/CFT will somewhat restrict the scope of my conclusions, I do this for two reasons:

(i) because the question of the status of the semiclassical limit of QG is inevitably a theory-specific matter, which requires a theory-specific assessment.

(ii) nonetheless, there is reason to believe that for the case of scattering amplitudes, the analysis that I give at least generalizes to other approaches.[2]

Indeed, I rely on a specific example from AdS/CFT for concreteness and to illustrate my construction with a specific example. I choose this example instead of ones which might appear more "realistic",[3] such as ones coming from string theory or loop quantum gravity, because AdS/CFT allows an especially precise description of both the relation between semiclassical

---

issues.

[2]See for example Huggett and Vistarini (2015) for a similar construction in the case of string theory and Wallace (2021) for one in low energy QG. Indeed, insofar as scattering amplitudes are the workhorse of modern particle physics, and we define semiclassical quantities in terms of them, the analysis of this chapter should generalize to any QG theory fitting a minimally quantum field theoretic treatment.

[3]Where by realistic here I mean that they could supply candidates to describe our world. AdS/CFT cannot, since our world is not AdS, but de Sitter. Though note that there are approaches being developed which extend AdS/CFT to cosmological, de Sitter spacetimes, and which preserve most of the features relevant to this paper (Balasubramanian et al., 2021; Chen et al., 2021; Bousso and Wildenhain, 2022; Bousso and Penington, 2022). However, to avoid unnecessary complications, I rely on the better understood and more developed context of AdS/CFT.

physics and fundamental QG, and the basic structure of the fundamental QG degrees of freedom since we can use the dual CFT to define a Hilbert space and operators for fundamental QG. As these features are harder to find in the aforementioned "realistic" examples,[4] AdS/CFT is especially suited to illustrate the basic features of the Humean position I articulate. Hence, while my conclusions are ultimately provisional on the validity of the relevant features crucial to these AdS/CFT constructions, there is reason to believe that the final theory of QG will save enough of that structure to make my discussion relevant.

Previous discussions of the disappearance of spacetime in QG as it relates to Humeanism include Matarese (2019); Wüthrich (2020); Jaksland (2021); Lam and Wüthrich (2021a). In particular, Jaksland (2021) provides what is arguably the main alternative to the approach described in this chapter, i.e., *entanglement fundamentalism*. This chapter provides an alternative viewpoint on Humeanism in QG to be evaluated in comparison with the literature mentioned above. Such comparison, however, will not be undertaken explicitly here and will be left for future work.[5] The reason behind this is

[4]For example, string theory is known almost exclusively perturbatively, with its non-perturbative formulation still mostly unknown (indeed, a source of interest in AdS/CFT and holography is as a possible route to define non-perturbative string theory). Loop quantum gravity, on the other hand, has problems in giving a fully worked-out dynamics for the theory.

[5]Though let me note that entanglement fundamentalism, while interesting and very promising, appears to depend very sensitively on the final theory of QG, relying in a very peculiar way on entanglement as the origin of spacetime and dynamics. It is unclear to what extent this assumption is borne out in various approaches to QG (though some certainly do). Indeed, see for example Giddings (2015); Susskind (2016b) for issues and limitations of the connection between entanglement and spacetime, and Lam and Wüthrich

that any such comparative evaluation would ultimately require a well-defined and empirically confirmed theory of QG to define which approach best fits its structure. Without such a theory, no comparative evaluation can be carried out in a truly satisfactory manner.

This observation also justifies the interest in a further proposal to reconcile Humeanism with QG. Lacking a final theory of QG, it seems reasonable, insofar as we want to do metaphysics based on actual QG theories, to explore as many options as possible for how this metaphysics might look like while leaving the choice of the best option to a time when the science will be more defined. Of course, any metaphysical approach developed in this way will be more or less theory-dependent and provisional and might end up misguided as we develop new theories and find new empirical data. I take this to be the price to be paid for the naturalistic commitment to do metaphysics based on our best science (Ladyman et al., 2007), coupled with the belief that certain theories of QG are well-defined enough to allow some non-trivial metaphysical reflection to be carried out on their basis.[6] Indeed, following Ladyman et al. (2007), one might take this dependence of metaphysics on empirical and scientific data yet to come to be the hallmark of naturalized metaphysics, and in this sense to be positive.

(2021a) for more metaphysical issues. These doubts would at least motivate the project of finding alternative approaches to Humeanism in QG, such as the one pursued here, in case entanglement is insufficient.

[6]A belief that seems to be shared by at least a non-trivial amount of authors, see for example Matarese (2019); Wüthrich (2020); Cinti and Sanchioni (2021b); Jaksland (2021); Lam and Wüthrich (2021a); Le Bihan (2020).

The chapter is structured as follows: in §4.1, I explain the issues QG raises for Humeanism and recast them in the language of causal structures. In §4.2, I propose to use amplitudes as the fundamental relation for the Humean supervenience basis, while in §4.3, we look at how this proposal functions and recovers semiclassical physics in a concrete example from AdS/CFT. §4.4 considers and replies to some objections, while §4.5 concludes.

4.1 Understanding Humeanism

This section aims to formulate Humeanism in a particularly suitable way for a generalization compatible with QG. For starters, we need a reasonably clear statement of the Humean thesis. A particularly influential one is provided by Maudlin (2007) and goes as follows:

> **Separability**: the complete physical state of the world is determined by (supervenes on) the intrinsic physical state of each spacetime point (or each pointlike object[7] and the spatiotemporal relations between those points.
>
> **Physical Statism**: all facts about the world, including modal and nomological facts, are determined by its total physical state.

In this definition, we can understand **physical statism** as encoding the basic tenet of Humeanism that modal matters can be reduced to occurrent ones,

[7]Note that here I use the word object as a sui generis term, without commitments to any particular metaphysics of objects.

typically realized via a best system analysis (Ramsey and Mellor, 1978; Lewis, 1973, 1983; Earman, 1984). In contrast, **separability** encodes the structure of the supervenience basis for said reduction. It is this second clause that will interest us the most.

The appeal to spatiotemporal relations in this context can be understood, following Lewis (1986), as being a consequence of spatiotemporal relations being the worldmaking relations. Since a worldmaking relation is supposed to unify and structure a world in a Humeanly acceptable way, it is then natural to expect them to give the appropriate supervenience basis for Humeanism. Coupling this statement with the observation that spatiotemporal relations are naturally understood as obtaining between pointlike entities, we get **separability**. Already Lewis recognized that spatiotemporal relations might prove too restrictive as a requirement on worlds,[8] and thus allows that *analogously spatiotemporal* relations are enough. Analogously spatiotemporal relations, for Lewis, are *natural, external, pervasive*, and *discriminating*.[9]

As we want to extend Humeanism to QG, it seems natural to expect that even analogously spatiotemporal relations will not be enough, for QG, insofar as it entails the complete disappearance of spacetime geometry (Huggett and Wüthrich, 2013), is presumably not even analogously spatiotemporal. Indeed, the need to move beyond analogously spatiotemporal relations was highlighted already in Darby (2009, 2020) for reasons of quantum entan-

[8]For example, how could we account for a Newtonian world where space and time are separate?

[9]See Lewis (1986) for the definitions of these terms)

glement, and Bricker (1996) for more metaphysical reasons related to the possibility of island universes. Finally, Wüthrich (2020) argues that, in the case of theories of QG where spacetime is at most an emergent entity, there are no analogously spatiotemporal relations available to act as worldmaking relations for QG.[10] I thus follow these authors in taking a relation being natural and external as a sufficient and necessary condition for it to be a worldmaking relation.

Let me now introduce a different notion that will be helpful in our quest to generalize Humeanism to QG. The notion I have in mind is that of a causal structure, already encountered in chapter 2[11] and defined as follows:

(CS) **Causal Structure**: given a theory T, we say that a causal structure of the theory T is given by a set of spacetime points/pointlike objects (with their intrinsic physical state) and a relation R which determines if two objects/points of spacetime can or cannot be causally related.

Recall from chapter 2 that when I speak of causation in **(CS)**, I only mean that the correlations determined by R are counterfactually stable. The use of the idiom of causation is mainly for ease of exposition, and a reader skeptical towards a broadly counterfactual analysis of this notion is welcome to substitute causality talk with talk of counterfactually stable correlations.

[10] Wüthrich (2020) also argues that, for the specific case of causal set theory, there is no natural external relation that acts as a worldmaking relation. The rest of this chapter can be seen as an argument that this conclusion does not apply to theories of QG where scattering amplitudes are defined.

[11] For more on the use of causal structures to understand QG, see Cinti and Sanchioni (2021b,a).

To grasp how **(CS)** works, we can return to our two basic examples: the causal structure of spacetime and that of entanglement. The causal structure of spacetime is a causal structure whose background theory is GR, whose objects are spacetime points[12], and whose generating relation R is the relation R_{LC} of *being connectable by a causal curve* obtaining between two spacetime points p, q, where a causal curve is a curve that is either timelike or null. Note that the causal structure of spacetime is sufficient to describe the basic structure of relativistic spacetime and, in particular, to encode all matters concerning locality and propagation of signals/matter.[13]

The second example is the causal structure of entanglement. This causal structure takes as its background theory quantum theory, as its objects quantum objects, i.e., objects whose state is a quantum state, and as its generating relation R the relation R_E of *being entangled* obtaining between two quantum objects.[14] For an argument that entanglement correlations entail appropriate counterfactuals, see Maudlin (1994). The causal structure of entanglement gives the basic structure of non-local (from the perspective of spacetime) correlations induced by a specific quantum theory. Note that

[12]I here appeal to spacetime points for ease of exposition, though the discussion given here can be straightforwardly extended to more sophisticated approaches to the ontology of spacetime.

[13]Indeed, modulo fixing a conformal factor, and under assumption that certain mild causality conditions are satisfied, everything about spacetime is fixed by what I call the causal structure of spacetime, thanks to a theorem of Malament (1977).

[14]Note that in the background of this definition there is a realist understanding of entanglement (Glick and Darby, 2018; Cinti et al., 2022). Anti-realists about entanglement might have issues identifying the causal structure of entanglement outside a purely operational understanding of it.

for this chapter, I will not take any specific stance regarding the ontology of quantum objects/systems, much less their ontology in the context of QG. Instead, I will take quantum objects simply as generic entities capable of supporting a representation as Hilbert space states or, possibly more generally, von Neumann algebras. Relatedly, I will treat subsystem relations in purely formal terms, without taking any stance on their specific ontological description, either as subspaces of Hilbert spaces or as subalgebras of von Neumann algebras.[15] A pointlike quantum object will be, for present purposes, any object represented by the subspace of the Hilbert space of the universe of smallest possible dimension (or smallest subalgebra of the von Neumann algebra of the universe).

An essential point for our purposes is that even for minimal combinations of quantum theory and GR, we will have a combination of the entanglement and spacetime causal structures. The details of how this combination takes place and which causal structure emerges from this process are, to some extent, dependent on the theory under consideration. In its simplest form, for example, in very straightforward semiclassical approaches to QG, one would superimpose the two causal structures. This procedure might, however, lead to paradoxes and inconsistencies, as indeed often happens in similar naive attempts at semiclassical gravity. The AMPS paradox, and its

[15] Though note that, to extend the discussion of this chapter to the case of wavefunction realism (Ney, 2021) one would also need a Humean story to connect the wavefunction to 4-dimensional spacetime. I take this issue, however, to be general to proponents of wavefunction realism and not specific to the present discussion.

attendant resolution as discussed in chapter 2 would be a classic example of this phenomenon. More sophisticated combinations are possible already in a semiclassical approximation, as evidenced, for example, by the spacetimes emerging from the ER=EPR conjecture (Maldacena and Susskind, 2013), and become inevitable when we move to full QG, where spacetime often does not make sense. The distinction between entanglement and spacetime causal structure becomes a relic of a bygone semiclassical era. It is then helpful to have a more flexible notion of causal structure, the *generalized causal structure* that we already employed to make sense of the AMPS paradox. The generalized causal structure is a causal structure given by a background theory of QG, a set of quantum objects as objects,[16] and as generating relation R the relation R_{WH}[17] of *being non trivially related to*, obtaining between two quantum objects whenever there is a counterfactually stable correlation between them.[18] The generalized causal structure, as defined, captures the physical content of an arbitrary theory of QG, with or without spacetime, insofar at least as inducing counterfactually stable connections can be taken as a minimal condition for something to count as part of the physical content of a theory, which seems a reasonable assumption at this point. Moreover, note

[16] Though remember my observation before on the largely formal nature of my treatment of quantum objects

[17] Here WH stands for wormhole, and references the wormholes of the ER=EPR conjecture. I use this notation following Cinti and Sanchioni (2021b,a) where the generalized causal structure was initially introduced, with explicit reference to ER=EPR-type constructions.

[18] At least, I take this condition to be necessary for a non-trivial connection. Possibly, there could be other conditions, owing to the requirement that the relevant connection is physically significant, though, for simplicity, I will only consider counterfactual stability.

that, insofar as the generalized causal structure captures any counterfactually stable correlation, there should also be an embedding of the spacetime and entanglement causal structures. I will return to this issue in §4.3.

With the generalized causal structure, we can reformulate the primary task facing the Humean vis-à-vis QG in a way more amenable to resolution. To understand this point, recall that in standard Humeanism, the structure of the supervenience base is encoded in the axiom of **separability**. Insofar as we restrict ourselves to relativistic spacetimes, we can straightforwardly reformulate the axiom of **separability** in terms of the causal structure of spacetime as follows:[19]

> **Separability**: the complete physical state of the world is determined by (supervenes on) the causal structure of spacetime.

For relativistic spacetimes, this reformulation is equivalent to the original one of **separability** because the causal structure of spacetime is defined in terms of spacetime points and their properties, as encoded by their physical state, together with the causally relevant spacetime relations, i.e., causal curves. Since the state of the world presumably cannot be influenced by non-causal, i.e., spacelike, curves in a relativistic spacetime, the relation R_{LC} and the

[19] Note that the restriction to relativistic spacetimes here is acceptable. A fully general reformulation of **separability** would involve a variety of causal structures of spacetime, each defined in terms of a different relation R defining causal connectedness for a certain class of spacetime geometries. For example, Galilean spacetime could have a causal structure defined in terms of the relation R_G of *being in spatial contact with*. However, insofar as we are interested in QG, we are interested in a theory whose classical limit is GR, on pains of empirical failure. Hence, when trying to formulate Humeanism in a way compatible with QG, it is natural to start from a reformulation adapted to GR and then to appropriately generalize it, as I do in the following.

associated causal structure of spacetime should be enough to account for the spacetime relations of interest in the original formulation of **separability**. Hence, unpacking the definition of spacetime causal structure in **separability** simply gives back Maudlin (2007)'s formulation of **separability**.

Given this new formulation of **separability** and our understanding of the generalized causal structure of spacetime, we can immediately formulate an extension of **separability** that can account for non-spatiotemporal, quantum gravitational worlds. We just need to define ***generalized separability*** as follows:

> **Generalized Separability**: the complete physical state of the world is determined by (supervenes on) the generalized causal structure.

The axiom of **generalized separability** in principle allows us to capture the essence of a Humean supervenience basis in QG. It has been argued, in Cinti and Sanchioni (2021b), that **generalized separability** is capable, at the semiclassical level, of reconciling Humeanism and entanglement, at least for worlds satisfying something like the ER=EPR conjecture. Nonetheless, it is limited in that it is defined in terms of the notion of non-trivial connection, whose physical meaning is at least unclear. Indeed, as defined, it is not clear that R_{WH} would actually correspond to a single natural external relation rather than a collection of them, or even to no discernible natural external relation. In other words, we need to specify an appropriate natural external relation which gives rise to R_{WH} and the generalized causal structure. This relation will then fulfil the role of structuring the supervenience basis for quantum gravitational Humeanism. Hence, the main merit of **generalized separability** is not that it resolves the issues of Humeanism in QG, but instead that it gives us a language to formulate the goal of a Humean analysis of QG worlds. We can then put the question of Humeanism in QG succinctly as follows:

(QH) Is there a natural external relation well-defined for a QG world and which gives rise to R_{WH} and the generalized causal structure, i.e., which satisfies **generalized separability**?[20]

my goal in the next section will be to propose such a relation.

4.2 Amplitudes and QG Humeanism

In this section, I aim to introduce a proposal for a quantum gravitational extension of the standard Humean thesis and discuss how it satisfies the basic desiderata laid out in the previous section, notably **(QH)**. I will propose to identify the fundamental relation structuring the supervenience basis for the Humean around the relation R_Q of *having non-zero scattering amplitude*, obtaining between two quantum objects. In particular, I use R_Q to define the *quantum causal structure*, given by a background theory of QG, with quantum objects as its objects and R_Q as its generating relation. We will then see that R_Q is a natural and external relation in QG, and that the

[20]One presumably would also like to know how to account for non-fundamental facts, i.e., anything less or equally fundamental as semiclassical gravity and the entanglement and spacetime causal structures. We will see how to do this in a concrete example in §4.3.

quantum causal structure provides an answer to **(QH)**.

First, however, let me clarify what I mean by scattering amplitude in this context. When we speak of scattering amplitudes, we mean an expression of the following sort:

$$\langle out|\mathbf{S}|in\rangle = n \in \mathbb{C} \tag{4.1}$$

the expectation value of the so-called ***S**-Matrix* between two Hilbert space states. The **S**-Matrix is an operator effectively describing scattering in the sense that its matrix elements are scattering amplitudes. The **S**-Matrix acts on a Hilbert space $\mathcal{H}$, which has two distinguished subspaces $\mathcal{H}_{in}$ and $\mathcal{H}_{out}$, describing the subsystems related by the amplitude. In particular, the two states are called $|in\rangle$ and $|out\rangle$ because they represent the initial and final state of a scattering process, i.e., a process by which certain quantum objects interact with each other, leading to a new set of objects.[21] The paradigmatic example of this process in particle physics are the interactions taking place, for example, at the LHC. Note that while we are now using temporally loaded terms like initial and final state, no reference to a background (spatio)temporal order is, in principle, required to make sense of this expression.

[21] Strictly speaking, the $|in\rangle$ and $|out\rangle$ states in (4.1) are asymptotic states, i.e., they are formally defined as being at time $t = \pm\infty$. This fact is usually understood as purely formal since, otherwise, no real-world experiment would count as a scattering experiment. If the reader is troubled by this, note that an equivalent definition of scattering amplitudes is as follows: $\langle out(t)|\mathbf{S}|in(t)\rangle = n \in \mathbb{C}$, where the states now have explicit temporal dependence and are not asymptotic. For ease of exposition, I will stick with the definition of (4.1) in what follows.

Indeed, scattering amplitudes, as defined here, are inner products between states. As we can define quantum states and their inner products, i.e., work in Hilbert space, we can have scattering amplitudes. Moreover, diffeomorphism-invariant schemes for defining scattering amplitudes have been defined. See Rovelli and Vidotto (2015) for an example in loop QG, and Witten (2018) for discussion of the issue in string theory.

We can compute scattering amplitudes through an appropriate path integral via the following formula:

$$\langle out|\mathcal{S}|in\rangle = \int \mathcal{D}\phi e^{-i\mathcal{S}[\phi]} \tag{4.2}$$

where $\mathcal{S}$ is the action of the underlying classical theory being quantized, and $\mathcal{D}\phi$ is the (in general ill-defined) measure on the space of fields ϕ.

Moreover, by the usual tools of QFT, most notably the LSZ formula (Srednicki, 2007), we can also express a scattering amplitude as a correlation function, i.e., an expression schematically of the following sort:

$$\langle out|\mathcal{S}|in\rangle \propto \langle\chi|\, T\phi(x)\ldots\phi(y)\ldots|\chi\rangle \tag{4.3}$$

where T is a time order operator, i.e., an operator ensuring that the ϕs are ordered from past to future, the state $|\chi\rangle$ can either be $|\Omega\rangle$, the vacuum state, or any other possible state of our theory, while $\phi(x)$ is a field operator, i.e., an operator acting on a quantum field at a point x in the manifold.

Intuitively, we can think of scattering amplitudes as giving us the amplitude that a certain initial state evolves into a certain final state. In general, states that can transform into each other have non-zero amplitude, while states that cannot evolve into one another have zero amplitude.[22]

Note that I here speak of amplitude and not probability since amplitudes, as usual in quantum theory, are not probabilities. In particular, to get a probability out of a scattering amplitude, we would need to take the modulus squared of the amplitude. Hence, scattering amplitudes do not straightforwardly have a modal character, at least not the modal character of probabilities, making them prima facie acceptable for the Humean. Indeed, Earman (2021) shows that starting from a quantum system described by operator algebras, states, and amplitudes, i.e., the structure of the generalized causal structure as defined here, quantum chances for such a system can be given a Humean analysis along the lines of Lewis (1994). Amplitudes then are not Humeanly unacceptable posits due to an intrinsic probabilistic nature since they are strictly speaking not probabilities; and while there certainly is a tight relation between amplitudes and probabilities, this connection should not be seen as a problem for the Humean, but as a resource since, following Earman (2021), we can see amplitudes as part of the supervenience basis for a Humean analysis of quantum probabilities, rather than as probability-like entities themselves. In a slogan, the Humean does not understand amplitudes starting from probabilities, but rather probabilities starting from amplitudes;

[22]Hence, in the proposal that I am going to develop in the following, they cannot be worldmates.

and since there is no a priori reason to think of amplitudes as modal once their connection with probabilities is understood in this way, amplitudes are acceptable for the Humean. I will say more about the modal character of amplitudes in §4.4.2.

Moreover, scattering amplitudes are well-defined for the major approaches to QG capable of dealing with a non-trivial notion of quantum evolution. In particular, both string theory (Green et al., 2012) and (covariant) loop QG (Rovelli and Vidotto, 2015) compute scattering amplitudes. This fact is also true of AdS/CFT (Ammon and Erdmenger, 2015) and low-energy QG (Wallace, 2021). In all these cases, we can think of the definition of the relevant scattering amplitudes in terms of inserting an appropriate action principle in a path integral-like expression.

After this lightning-quick introduction to scattering amplitudes, let us move to the core of this section. My main proposal will be to regard scattering amplitudes, formally encoded in the way described above, as arising from an actually existing scattering amplitude relation obtaining between actually existing quantum systems. This proposal is supposed to be analogous to how metric relations encoded by a metric on a manifold encode actual spacetime relations instantiated among actually existing entities. To complete my argument that these scattering relations are sufficient for the Humean project, as understood here, I need to show that they are natural external relations answering **(QH)**.

Naturalness. Let me start from naturalness. By naturalness, I mean that the relations encoded by scattering amplitudes are part of the basic vocabulary best suited to express the fundamental content of the world.[23] Insofar as we take QG to encode the basic structure of the world, we can speak of the content of the relevant QG theory under consideration. In this way, we can immediately see that scattering relations are natural in the relevant sense, since they constitute arguably the most basic quantity that can be associated with a given quantum theory. Indeed, insofar as the whole set of scattering amplitudes describing a certain situation fixes a specific path integral (4.2), and defining a specific path integral is the same as specifying a certain quantum theory, then specifying this set of scattering amplitudes is the same as specifying a certain quantum theory. Hence, scattering amplitudes should count among the natural relations of a given quantum theory where they exist. Given that they exist for the major approaches to QG, they should count as promising candidates for a relation structuring the supervenience basis.

Externality. We can now look at whether scattering relations count as external relations. By an external relation, I will mean satisfying the following definition, given by Bricker (1996):

> An external relation is one that, although it fails to supervene on the intrinsic natures of its relata, does supervene on the intrinsic natures of its relata and of the fusion of its relata.

[23] See Lewis (1983) for the original notion, and Dorr and Hawthorne (2013) for a detailed discussion of its definition.

Here, 'intrinsic' means a property preserved among duplicates of a given entity. To check that a scattering amplitude satisfies this definition, we first need to know what it means to take the mereological sum of two quantum systems. Since it is difficult to specify a mereological system adapted to quantum physics, we will limit ourselves to the assumption that, given two quantum systems, their mereological sum is at a minimum given by a Hilbert space having the quantum systems' Hilbert spaces as subspaces. I justify this claim because given my definition of subsystems as Hilbert space subspaces, then the Hilbert space encompassing said subspaces will have states specifying the properties of both subsystems, as expected of a mereological sum. Moreover, let us take the intrinsic nature of a quantum system to be specified by its quantum state, since that state will specify the expectation value of all observables for the system. We can now check that scattering amplitudes are external relations. First, we check that a scattering amplitude among two quantum systems cannot be defined via their intrinsic natures alone. To see this, one needs to see that scattering amplitudes are computed via the **S**-Matrix defined on a space $\mathcal{H}$ having $\mathcal{H}_{in}$ and $\mathcal{H}_{out}$ as subspaces (4.1), which means that scattering amplitudes cannot be defined only in terms of the Hilbert spaces of the two subsystems.[24] Rather, we should take into

[24] One might worry that $\mathcal{H}_{in}$ and $\mathcal{H}_{out}$ are not two different subsystems, but rather one the temporal evolution of the other. Note however that this distinction, in the present context, is irrelevant, since Humeans are interested in relations obtaining between pointlike objects. Hence, temporal separation, for the Humean, implies that we are dealing with two distinct entities between which an appropriate external relation should obtain (for example, a temporal relation, or an amplitude relation as in the present context).

account the whole space $\mathcal{H}$, i.e., their mereological sum by the stipulation above, so that we can define an **S**-Matrix element between the $|in\rangle$ system and the $|out\rangle$ system. However, this **S**-Matrix element inevitably depends on combining both systems and their intrinsic natures, i.e., on their mereological sum, as evidenced by the fact that we are using the Hilbert space $\mathcal{H}$. Thus, scattering amplitudes are external relations.

(QH). Let me now conclude my argument that scattering amplitudes constitute a suitable candidate for the fundamental relation structuring the supervenience basis of the Humean by checking that they satisfy **(QH)**. To do this, we need to show that the quantum causal structure defined by scattering amplitudes can account for the data of the generalized causal structure. In other words, the relation R_Q of having non-zero scattering amplitude should encode all non-trivial, i.e., counterfactually stable, connections between quantum systems, as given by R_{WH}. To see this point, recall that we can organize scattering amplitudes inside the **S**-Matrix, i.e., an operator whose matrix elements are scattering amplitudes. The **S**-Matrix is supposed to encode, for an arbitrary QFT, all the empirical information about that theory. If we were able to extend this claim about empirical content to one about the physical relations admitted by the theory, we would have answered **(QH)** positively, insofar as giving rise to counterfactually stable connections can be reasonably seen as a necessary condition for a relation to qualify as physical.[25]

[25] One might worry that counterfactual stability being only necessary and not sufficient might lead to problems. Recall, however (§4.1), that any other condition on physical

Indeed, this hope has given rise to a program to axiomatize QFT in terms of **S**-matrices.[26] While this program ultimately failed in QFT, in favor of the now standard Lagrangian approach, it is interesting to reconsider the original idea behind it in the context of QG. Note that I do not wish to defend anything analogous to the **S**-Matrix program for QFT in QG, but rather that it is reasonable to expect that QG, in particular, should be friendly to an account of its physical relations entirely in terms of the **S**-Matrix.

Note first of all that, already in QFT, the main obstacle to the **S**-Matrix program was not necessarily that **S**-Matrix did not encode any specific physical fact about QFT, but rather that it was hard to identify sufficient and necessary conditions for an **S**-Matrix to give rise to a QFT. In other words, what was lacking was an axiom system for QFT purely in terms of the **S**-Matrix. However, as is clear from (4.1), computing **S**-Matrix elements is equivalent to defining a path integral or a set of correlation functions. Insofar as those two objects are sufficient to account for all physically significant relations in a QFT, which is arguably the received view in particle physics, then the **S**-Matrix accounts for its physical content too.[27]

Does this argument extend to QG? Not only does it extend, since the same definitions, at least formally, can be exported to QG. It arguably gets stronger. Let me spell out why. As I mentioned above, there are three prima facie math-

---

significance would give rise to an analogous restriction on R_{WH}, thus ensuring that the two causal structures still match.

[26]See Cushing (1990) for its history.

[27]One might here worry about entanglement and its relation to these notions. I will discuss this point in §4.4.1.

ematically equivalent ways to spell out the physically significant relations of a QFT: the **S**-Matrix, path integrals, and correlation functions. Are all three of these still physically significant in QG? Arguably, only the **S**-Matrix is.

Starting with the path integral, the main issue with this approach is that the path integral, as defined, is a purely formal expression since the measure on the space of fields it uses is not well-defined. The usual approach in particle physics is to sidestep this issue by evaluating the path integral in perturbation theory. Since the Feynman diagrams around which perturbation theory is structured can be derived from the path integral, this approach works.

However, especially in gravity, it is the non-perturbative sector of the theory, not captured by this approach, that is crucial since GR is not perturbatively renormalizable and hence has a well-defined perturbation theory only at very low energies. Thus, it seems that the gravitational path integral fails exactly in the high energy, non-perturbative regime where QG effects become large.[28]

When it comes to correlation functions, the issue stems from the diffeomorphism invariance of GR, which makes field operators like $\phi(x)$ problematic since they are, in general, not diffeomorphism invariant (Tong, 2009).[29]

It then seems that the **S**-Matrix is the best way to encode the physically significant relations appearing in QG, leaving path integrals and correlations functions as tools to express properties of scattering amplitudes rather than

[28]Specific approaches to QG might, in principle, be able to give rise to well-defined path integrals. In this case, however, their outputs would be equivalent to the **S**-Matrix since they would compute **S**-Matrix elements.

[29]In this case too, specific approaches might be able to sidestep this issue. An example is AdS/CFT, where correlation functions can be defined in the boundary CFT.

independent perspectives on QG. Consequently, the quantum causal structure defined via R_Q gives a positive answer to **(QH)**.

We thus see that R_Q is a natural, external relation providing an answer to **(QH)**. Thus, the quantum causal structure defined via R_Q provides a natural candidate for a Humean supervenience basis suitable to QG worlds. Indeed, we can summarize this proposal, using the reformulation in terms of causal structures, in the exchange of **generalized separability** with the following axiom of **amplitude separability**:

> **Amplitude Separability**: the complete physical state of the world is determined by (supervenes on) the quantum causal structure.

4.3 Semiclassical Physics from R_Q

In this section, I look at how semiclassical physics, i.e., the spacetime and entanglement causal structures, are recovered from the quantum causal structure in a concrete example in AdS/CFT. Once we recover the spacetime and entanglement causal structures, we can account for any other less fundamental fact via the usual Humean means since the entanglement and spacetime causal structures allow us to define a notion of **separability** analogous to that defined in §4.1,[30] and thus capable of accounting for all less fundamental facts. Therefore, this step is crucial to my proposal. We will look at the

[30] Modulo some procedure to account for entanglement, either, for example, by adding a new entanglement relation (Darby, 2020), or through ER=EPR (Cinti and Sanchioni, 2021b).

holographic description of a black hole generated from gravitational collapse. In particular, we employ Papadodimas and Raju (2013)'s state-dependent construction of the interior.

First, let us start with the Penrose diagram of a black hole obtained from gravitational collapse. Such a system represents a one-sided black hole.[31] In AdS/CFT, scattering amplitudes are computed most naturally via correlation functions, using something like (4.3), since the boundary CFT makes standard QFT techniques like this one particularly useful. To evaluate a correlation function for a semiclassical quantity in the bulk AdS, which is usually given by a bulk mode, in the CFT, we need to know the CFT field operator corresponding to said bulk mode. Insofar as we know how to do this, we know how to recover semiclassical physics, and thus the entanglement and spacetime causal structures, from the quantum causal structures. The standard way to find a representation of a bulk mode in terms of CFT operators (the so-called HKLL reconstruction)[32] relies exactly on the possibility of evolving via the equations of motion any bulk mode to the boundary. The result of this evolution out to the boundary is an operator in the dual CFT. However, when we are dealing with one-sided black holes, such a procedure fails since, for the right moving modes in the interior, there is no way to evolve them to the boundary with the bulk semiclassical equations of motion since they

[31] By a one-sided black hole I mean a physical system with one event horizon. As such, a one-sided black hole has only one exterior region.

[32] For a review of the various issues and techniques connected to representing bulk quantities in the boundary CFT, see Harlow (2018)

either meet the singularity or interact at high energies with the collapsing matter which formed the black hole. Thus, we seem unable to find a CFT representation for modes living in the interior of the black hole.

Papadodimas and Raju (2013) solved this problem starting from the following standard representation for exterior operators:

$$\phi_{CFT}(t, \mathbf{x}, z) = \int_{\omega>0} \frac{d\omega d^{d-1}\mathbf{k}}{(2\pi)^d}[\mathcal{O}_{\omega,\mathbf{k}} f_{\omega,\mathbf{k}}(t, \mathbf{x}, z) + h.c.] \tag{4.4}$$

Here, $\mathcal{O}_{\omega,\mathbf{k}}$ is a mode in momentum space of the CFT field $\mathcal{O}(t, \mathbf{x})$. t and $\mathbf{x}$ are, respectively, the time and space coordinates of the CFT, while z is an auxiliary coordinate from the perspective of the CFT, which we can intuitively think of as the radial coordinate of AdS. The function f is a mode function necessary to ensure that the operator so defined correctly represents the AdS quantity that we are trying to study.

They then proposed that we represent all modes, including those behind the horizon, with operators with the following form:

$$\phi_{CFT}(t, \mathbf{x}, z) = \int_{\omega>0} \frac{d\omega d^{d-1}\mathbf{k}}{(2\pi)^d}[\mathcal{O}_{\omega,\mathbf{k}}\ g^{(1)}_{\omega,\mathbf{k}}(t, \mathbf{x}, z) + \tilde{\mathcal{O}}_{\omega,\mathbf{k}}\ g^{(2)}_{\omega,\mathbf{k}}(t, \mathbf{x}, z) + h.c.] \tag{4.5}$$

Here, again, $g^{(1)}$ and $g^{(2)}$ are mode functions required to ensure locality of ϕ_{CFT}, while we interpret the rest of the equation the same as (4.4). In (4.4) and (4.5), the operators $\mathcal{O}$[33] are obtained by the standard HKLL procedure

[33] From here on, I suppress the indices ω and $\mathbf{k}$ indicating the frequency and momentum of the mode, as they are not necessary for our discussion.

of evolving a mode to the boundary of AdS spacetime, and we can think of them as representing exterior modes. On the other hand, the operators $\tilde{\mathcal{O}}$ are best understood as giving a CFT representation of the interior, right-moving problematic modes. In particular, they satisfy the following conditions:

$$\tilde{\mathcal{O}}\,|\psi\rangle = C\mathcal{O}^{\dagger}C^{-1}\,|\psi\rangle \tag{4.6}$$

$$\tilde{\mathcal{O}}\mathcal{O}\,|\psi\rangle = \mathcal{O}\tilde{\mathcal{O}}\,|\psi\rangle \tag{4.7}$$

Here, C is an invertible matrix, which guarantees that $\tilde{\mathcal{O}}$ acts on the state $|\psi\rangle$ (which, by the principles of AdS/CFT, encodes the bulk geometry) as interior operators would. (4.7) guarantees that $\tilde{\mathcal{O}}$ and $\mathcal{O}$ commute, as we expect them to since they represent interior and exterior modes, which are spacelike related. Ultimately, (4.6) and (4.7) tell us that the operators $\tilde{\mathcal{O}}$ give us a way to represent the interior of our one-sided black hole.

How were Papadodimas and Raju able to evade the problems connected to representing the interior of a one-sided black hole in a CFT, thus giving the reconstruction (4.5)? The core idea is that we should not expect the operators $\tilde{\mathcal{O}}$ to satisfy (4.6) and (4.7) in the entire Hilbert space of the CFT (which encodes the full non-perturbative definition of the QG theory) but instead only in a subspace of the CFT Hilbert space, encoding the measurements easily realized by a low energy observer. Note that this is the only regime where the naive semiclassical picture of smooth and local spacetime makes sense. We call this subspace the *little Hilbert space*. It is the little Hilbert space which is responsible for the state dependence that names the construction under study. The operators representing measurements available to a low energy observer are polynomials in the $\mathcal{O}$s and their conjugates with conjugation matrix C (see equation (4.6)), with the caveat that the degree of the polynomials is not too large[34] and that the energy of any operator in the little Hilbert space is not too large.[35] Let us call this set of operators $\mathcal{A}$.[36] Once we know which are the operators relevant to our construction, we then need to pick a state $|\psi\rangle$[37] in the CFT Hilbert space, satisfying a particular equilibrium condition,[38] which guarantees that this state will be dual to a black hole with a smooth horizon geometry. The little Hilbert space is the vector space formed by applying the operators in $\mathcal{A}$ to $|\psi\rangle$. Intuitively, we can think of it as the (holographic dual of the) space of semiclassical, effective field theory perturbations around the bulk geometry described by $|\psi\rangle$.

We can now understand the reconstruction of the interior of the black hole. We only expect $\tilde{\mathcal{O}}$ to satisfy (4.6) and (4.7) within the little Hilbert space since we need to assume these conditions only in the semiclassical, effective field theory regime. Thus, no operators represent the interior on the *entire*

[34]Since this would be equivalent to a measurement with a large number of operator insertions, which would take us out of the semiclassical, effective field theory regime.

[35]In particular, it must be smaller than the energy necessary to generate a black hole, as this would again take us out of the semiclassical, effective field theory regime.

[36]Note that $\mathcal{A}$ does not form an operator algebra since it is not closed under multiplication. As such, the little Hilbert space is not a Hilbert space but only a vector space.

[37]It is not an accident that we call this state $|\psi\rangle$. Indeed, it is the same state appearing in (4.6) and (4.7).

[38]There are various ways to specify this condition, see Harlow (2014) for a critical discussion of some options.

CFT Hilbert space, but only on the little Hilbert space of effective field theory quantities. Indeed, the $\tilde{\mathcal{O}}$s are state-dependent, since they are defined by their action on the state $|\psi\rangle$ from which we construct the little Hilbert space. This fact is evident from (4.6) and (4.7). Thus, for different equilibrium states (equivalently, different little Hilbert spaces), we have different representations of the interior operators.

What happens outside the little Hilbert space? The reconstruction procedure breaks down, and thus the semiclassical approximation is not reliable anymore. Equivalently, outside the little Hilbert space, there is QG. The breakdown of the reconstruction procedure and the semiclassical approximation entails that we are not allowed anymore to assume (4.7), i.e., that the interior and the exterior of the black hole commute and are thus distinct systems. Indeed, outside the little Hilbert space, $\left[\mathcal{O}, \tilde{\mathcal{O}}\right] \neq 0$. Moreover, we can reconstruct any bulk field in the black hole's interior by combining $\mathcal{O}$ and $\tilde{\mathcal{O}}$. Thus, it follows that it is not guaranteed anymore that a bulk field in the interior and one in the exterior commute. Consequently, the interior and exterior of the black hole are not independent systems, despite being, at first glance, distant, non-overlapping regions in spacetime. We can think of this as a prototypical QG phenomenon, which means that we need the full power of the quantum causal structure and R_Q to understand it. Indeed, we do since, by definition, such highly non-local connections cannot be understood in purely spatiotemporal terms, and the non-commutativity of $\mathcal{O}, \tilde{\mathcal{O}}$

implies that entanglement is not relevant either.[39] Rather, all we have at our disposal to understand this situation are the correlation functions, and thus, as argued in §4.2, ultimately the scattering amplitudes R_Q, defined for the holographic one-sided black hole, and nothing else. All we can rely on is the quantum causal structure.

In particular, from the perspective of causal structures, we identify the semi-classically non-local connections which appear outside the little Hilbert space with instances of the relation R_Q (which defines the quantum causal structure) that do not correspond to either a spacetime relation R_{LC} or an entanglement relation R_{ME}. Indeed, the embedding of the causal structures of spacetime and entanglement in the generalized causal structure discussed in §4.1, which extends to an embedding in the quantum causal structure (since R_Q gives rise to R_{WH}, as we saw in §4.2) is an isomorphism only in the semi-classical limit, or equivalently inside the little Hilbert space. In other words, only inside the little Hilbert space we can think of R_{LC} and R_{ME} as two fully distinct relations whose causal structures are superimposed. In particular, this isomorphism means that from R_Q, we can recover R_{LC} and R_{ME}, i.e., semiclassical gravity. Moreover, it also means that **separability** holds inside the little Hilbert space, possibly supplemented with an entanglement relation (Darby, 2020) or with ER=EPR-like connections (Cinti and Sanchioni,

[39] See Earman (2015) for discussion of the need for commutativity between subsystems to define entanglement. Note that the difficulties with defining entanglement make these types of constructions especially hard to fit into the entanglement fundamentalism framework of Jaksland (2021).

2021b). Therefore, from R_{LC} and R_{ME}, we can recover in a Humeanly acceptable way anything that is less fundamental than semiclassical gravity, i.e., everything about our world except for QG. Since, as we have seen, we can recover semiclassical gravity from the quantum causal structure, then from it, we can recover everything about our world in a Humeanly acceptable way. Thus, in the concrete case of AdS/CFT described here, the quantum causal structure and R_Q give a fully Humean analysis of a QG world. Indeed, outside the little Hilbert space, the causal structures of entanglement and spacetime break down, and with them, the semiclassical picture they describe. We are left with only the quantum causal structure, with its new connections, encoded via R_Q. These new connections make it the case that systems which, from the perspective of the causal structures of spacetime and entanglement, are distinct, separate systems, are interdependent and non-trivially connected, in paradigmatically QG fashion.

4.4 Objections

Having seen how scattering amplitudes can serve as a structuring relation for the Humean's supervenience basis, let us now turn to some possible objections one might have to this proposal. In particular, we will first look at how this proposal deals with entangled states (§4.4.1), and then I will discuss how scattering amplitudes do not imply a restriction on free recombination

(§4.4.2).

4.4.1 Entanglement and Amplitudes

The first criticism that one might have towards the proposal of using scattering amplitudes to construct the supervenience basis is that one cannot account for entangled states in this way. In particular, the worry would be that given that entangled states imply a failure of **separability** in the spacetime setting (Maudlin, 2007), which can be avoided, for example, by adding entanglement relations to the supervenience basis (Darby, 2020), a similar move would be required in the context of scattering amplitudes. More in detail, entanglement implies that quantum states cannot factorize as product states of their subsystem components. Moreover, quantum states are required to define scattering amplitudes. Then, one might worry that there must be both entanglement and scattering amplitude relations to construct the supervenience base. The reason for this worry would be that it seems prima facie complicated to account for entangled states (and scattering amplitudes between them) only in terms of their subsystems' states and scattering amplitudes between them. Insofar as the state of the world can be entangled, then either one countenances entanglement relations, which would imply a significant loss of simplicity for the scattering amplitude proposal, or one faces a failure of **separability**.

At the same time, however, we have seen in §4.3 that the causal structure of entanglement can be recovered from the quantum causal structure, which suggests that entangled states can indeed be accounted for in terms of scattering amplitudes. Indeed, we can. Let us briefly see how, working for simplicity with an interacting QFT in mind.[40] Take two field operators $\phi(x)$ and $\phi(y)$, which we can think of as operators exciting the quantum field at the spacetime point appearing in their argument. Moreover, assume that the two spacetime points x, y are at spacelike separation. We can think of these two operators as describing two systems, X and Y, associated with different regions of the quantum field at spacelike separation. In particular, acting on the vacuum Ω with $\phi(x)$ will give the state of system X, and acting on the vacuum with $\phi(y)$ will give the state of Y. Then, we can check whether the two systems are entangled by computing the following correlation function, which is related to an appropriate scattering amplitude by (4.3):

$$\langle\Omega|\, \phi(x)\phi(y)\, |\Omega\rangle \tag{4.8}$$

The state of the composite system of X and Y is product, i.e., not entangled, if (4.8) is 0, and entangled otherwise. Indeed, one can also, through a more complex amplitude computation, also quantify the amount of entanglement between the two systems via one of the various entropic measures associated with entanglement. For details on how to do this, see, for example, Harlow (2016).

Hence, there is no issue accounting for entangled states via scattering ampli-

[40]There is no substantial difference between this case and the QG case, at least insofar as the relevant formal expressions are well-defined in QG, as one expects.

tudes and quantum states associated with the subsystems. Thus, there is no need to add entanglement relations to the supervenience base, and scattering amplitudes are sufficient for the Humean.

4.4.2 Amplitudes and Free Recombination

Let us now see whether my proposal is compatible with the principle of free recombination, which roughly asserts that any two distinct entities coexist in some possible world (Lewis, 1986). The principle of free recombination is another way of encoding the Humean prohibition against necessary connections among distinct entities, and thus plays a crucial role in (most versions of) the Humean project. The core of this worry is expressed well by Lam and Wüthrich (2021a), who, while discussing a proposal somewhat similar to the present one in the context of LQG, note that there is a conflict between free recombination and the use of concrete physical theory to identify the relation structuring the supervenience basis, as is the case for the present amplitudes proposal. This conflict is supposed to stem from the fact that things like scattering amplitudes are too closely connected to physical laws and thus inevitably carry a degree of *nomic voltage* (Lam and Wüthrich (2021a)'s words), which makes them unsuitable for the Humean. In particular, if amplitudes are so closely connected to laws of nature, then there seems to be a conflict with free recombination since, by free recombination, we should be able to recombine distinct entities as we want, while laws of nature presumably would forbid certain combinations. If amplitudes are laws of nature in disguise or nonetheless carry ineliminable nomic voltage, then free recombination is in danger.

As a first answer, let me note that there seem to be three different ways to understand this objection:

(i). *Amplitudes are laws of nature or, nonetheless, require laws of nature to exist.* Note that, prima facie, this argument seems dead in its tracks: amplitudes are external relations (§4.2) and do not have the required modal force to be laws of nature. At least, any argument of this form should include an argument that amplitudes are not external relations. One might argue for this claim, somewhat indirectly, from a formula like (4.2), where we express amplitudes in terms of a path integral with a relevant action principle. If we can take path integrals and action principles as laws of nature, then it seems that amplitudes do indeed have modal strength and thus cannot be external relations.

Does this argument show that amplitudes are too tightly connected to laws of nature, and thus unsuitable for the Humean? It seems not. In particular, all (4.2) shows is that, given a suitable set of scattering amplitudes, we can determine the relevant laws of the theory as given by its path integral. However, this fact is just the restatement of the claim that amplitudes are sufficient to fix the nomological facts at a world, not that they depend on such facts. In particular, amplitudes are, formally, nothing more than the inner product on the Hilbert space of the theory. While a specific inner product, i.e., a specific combination of amplitudes, is indeed equivalent to a specific path integral, the general existence of a path integral, i.e., the existence of some scattering amplitudes, independently of the choice of a precise combination, is nothing more than the claim that the space of states is a Hilbert space.[41] Such a claim seems straightforwardly acceptable for the Humean. It might be problematic insofar as one thinks that quantum objects are identical to Hilbert space vectors, i.e., a form of wavefunction realism, thus making *being a Hilbert space* a condition on free recombination, but such an assumption is not required. Insofar as Hilbert space is simply a formal tool for the representation of quantum objects, its status seems to be no different from that of classical state spaces, whose appearance in the mathematical representation of classical physics is compatible with Humeanism.

(ii). *There could be worlds where scattering amplitudes are not the relevant worldmaking/supervenience-basis-structuring relation.* Thus, scattering amplitudes would place an undue restriction on free recombination. To assuage this worry, simply note that there is no need for such an assumption. In particular, following Bricker (1996); Darby (2009); Wüthrich (2020), all that is required for a relation to be a candidate worldmaking/supervenience-basis-structuring relation is that it is a natural external relation. At no point does one need to assume that this relation is the same at all worlds.[42] Indeed, while

[41]Since inner products distinguish Hilbert spaces from normal vector spaces.

[42]At least under a minimal assumption of naturalism, as I am making here. If one were to accept the full scope of Lewis (1986)'s project of having a pre-theoretic way to understand modality in a Humean way, it might be problematic to allow such variation in what counts as worldmaking. In particular, one might worry that worlds are unified by relations alien to our pre-theoretic conception of the world (indeed, scattering amplitudes would fit this description somewhat, since they presumably are not part of our pre-theoretic

endorsing spatiotemporal relations, Lewis himself noticed a similar problem and proposed to move to analogously spatiotemporal relations, which include spatial and temporal relations. The move I suggest is the same. Natural external relations structure worlds. Which natural external relations should be used is a matter to be decided on a world-by-world basis, starting from appropriate empirical/scientific considerations. I propose that *for the particular case of QG worlds*, scattering amplitudes are a natural candidate relation. Which relation should be used for non-QG worlds is not touched by the present discussion.

(iii). *There seems to be a general conflict between Humeanism and a general form of naturalism, which says that we should look at laws for guidance in metaphysics.* As the Humean is committed to the non-fundamentality of laws of nature, it seems that the naturalist's commitment to the role of science in metaphysics is problematic. In other words, why should scientific theories be guides for metaphysics if scientific theories are just useful summaries of occurrent facts? This worry certainly points to an interesting issue for the Humean. This issue, however, is not specific to amplitudes or even to QG in general. Rather, it is a worry about Humeanism in general and its naturalistic credentials. As such, it goes beyond the scope of the

worldview). I think that on this point, a naturalistically motivated Humean, as I am, should simply deny that such pre-theoretic considerations should play a significant role in motivating our metaphysics. Indeed, whether the reduction of modal facts to non-modal facts can be carried out pre-theoretically, while important for Lewis's project, seems to go beyond the basic commitments of the Humean, as summarized in §4.1; as such, it is not a pressing concern for my proposal.

present proposal, which assumes at least the general viability of the Humean project before any QG concerns. Nonetheless, a possible answer may come from the observation above that Humeans are entitled to use scientific theory to determine the appropriate natural external relations and occurrent facts describing our world.[43] In this sense, the Humean is perfectly naturalist. Where the Humean seems to diverge with the naturalist, at least as understood in Lam and Wüthrich (2021a), is in not thinking that scientific theories are also guides to the *modal nature* of reality. However, whether this commitment is necessary for naturalism seems an open question, as evidenced by the contingent of Humean naturalists/scientific realists (Earman, 1984; Lewis, 1994; Loewer, 1996; Psillos, 2014; Cohen and Callender, 2009).

4.5 Conclusions

We have seen a possible route based on scattering amplitudes to reconcile Humeanism with the lack of spacetime that QG is expected to show. Whether this proposal is correct will ultimately depend on the specifics of the final theory of QG. In the meantime, it seems reasonable to explore all possible alternatives, and the present proposal is made in this spirit.

[43] As in footnote 42, there might be an issue if one couples Humeanism with Lewis (1986)'s project of specifying the pluriverse pre-theoretically. However, I take it that this coupling with the full Lewisian metaphysical project is not necessary for Humeanism in general, and indeed naturalistically motivated Humeans should probably abandon such a priori projects. For a version of Humeanism that does away with Lewisian modal realism, and more generally with a priori, non-science-based metaphysical theorizing, see for example Norton (2022).

In the next chapter, I will focus on how spacetime emergence is realized in models like the one discussed here. I do this for two reasons: first, because this issue is crucial to understand the metaphysics underlying these QG models; second of all, because understanding the functioning of spacetime emergence in this context is important in general to understand how to treat non-fundamental laws, which in QG includes GR and the standard model of particle physics, and hence to understand how Humeanism as realized in these models relates to the non-fundamental world and hence how everything can supervene on the basic QG degrees of freedom. Indeed, the discussion of the next two chapters will allow me to greatly improve the description given here of how semiclassical physics is recovered from AdS/CFT, and hence of how Humeanism in QG actually functions.

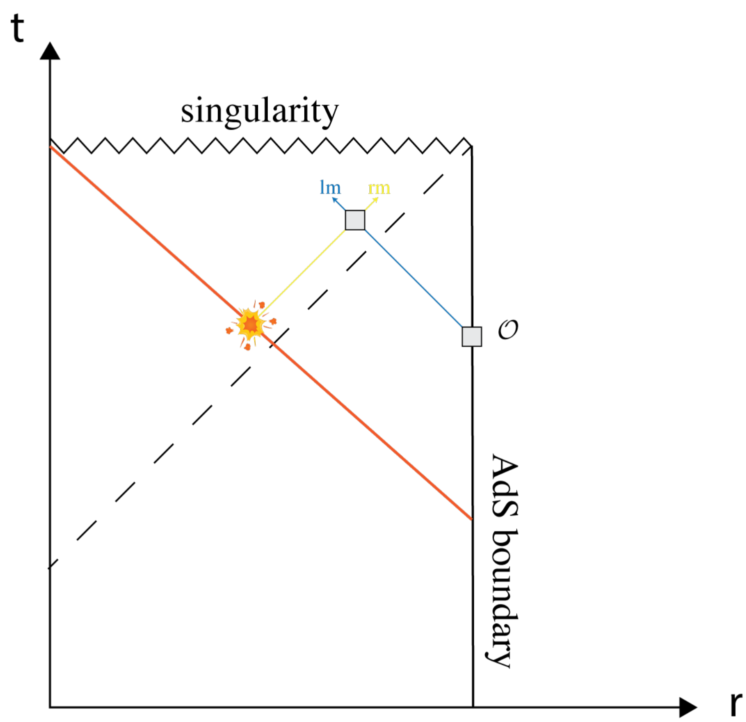

Figure 4.1: The Penrose diagram of a one-sided black hole formed from gravitational collapse. The dual CFT sits at the boundary of AdS, on the left of the diagram. The orange line represents a collapsing shell of matter which generates the black hole (note that a shell of matter collapsing at the speed of light as in the picture is an idealization; I employ it because it makes it easier to visualize various claims made in the main text). I have also highlighted an interior degree of freedom with its right-moving rm (yellow) and left-moving lm (blue) modes. The left moving part, which behaves like an exterior mode, can be evolved back to the boundary with the semiclassical bulk equation of motion, where $\mathcal{O}$ represents it. On the other hand, the right-moving mode, when evolved back to the boundary, crashes with the collapsing shell of matter at high energy, rendering the semiclassical bulk equation of motion useless (explosion).

Chapter 5

The Fate of Spacetime in Holography

We now move in this chapter to consider how spacetime emergence is realized, if at all, in the AdS/CFT models of quantum black that I have discussed in previous chapters. Indeed, understanding how to characterize the emergence of spacetime is one of the main philosophical problems connected to developing a theory of QG. In particular, the philosophical literature identifies two main problems connected to the emergence of spacetime: the empirical incoherence (Huggett and Wüthrich, 2013) and the hard problem of spacetime (Le Bihan, 2021). The problem of empirical incoherence concerns the meaning and possibility of measurements, understood as events localized in space and time, in a theory without spacetime. Following Huggett and Wüthrich (2013), the usual answer is that what is crucial is that spacetime is defined at least in an emergent sense, even if it is not there fundamentally. The hard problem of spacetime concerns, roughly, the definition of a relation of emergence explaining how spatiotemporal concepts can emerge (be characterized) from the non-spatiotemporal fundamental concepts of QG. To get a feel for some positions in the debate, let me briefly summarize some relevant options without any pretence of being exhaustive. Lam and Wüthrich (2018, 2021a) argue that in both cases, a form of functionalism about spacetime, which we can characterize for now via the slogan "spacetime is as spacetime does", provides a resolution to these problems. The problem of empirical incoherence is resolved by observing that GR must emerge from QG in an appropriate limit and that in that limit, spacetime will be emergent in the sense that the fundamental QG degrees of freedom will behave like spacetime and thus, according to functionalism, be spacetime. The hard problem of spacetime is solved by Lam and Wüthrich (2018) by observing that, strictly speaking, there is no hard problem of spacetime once we accept the functionalist point of view: once we have identified an entity behaving appropriately, that entity will be spacetime and thus fall under the appropriate spacetime concepts. In both cases, the crucial aspect of the functionalist approach is that we can define, in QG, an appropriate structure which behaves in the same way as spacetime. Baron and Le Bihan (2022a,b) opt instead[1] for a non-spatiotemporal mereology, whereby spacetime is composed of non-spatiotemporal building

[1]Though note that in principle these options do not have to be exclusive. It would be interesting to explore the possible combinations of these approaches, though it goes beyond the scope of the present work.

blocks. Again, empirical incoherence is avoided because spacetime is recovered as an emergent, composed entity, while the hard problem of spacetime is solved via appeal to mereology as the relevant relation. The last example, Baron (2021), suggests a form of spacetime eliminativism, where spacetime is taken not to exist. A helpful analogy here is with Minkowski and Galilean spacetime, where it is usually taken not to be the case that the latter emerges from the former, but rather that there is only Minkowski spacetime and that in the appropriate regime of low velocities, it approximates Galilean spacetime. The same is supposed to happen between the fundamental QG degrees of freedom and GR spacetime. In this case, the hard problem rather than avoided is voided: much like eliminativist approaches in the philosophy of mind (Ramsey, 2022), there is simply no hard problem because there is no mind/spacetime. Empirical incoherence is somewhat more difficult to solve for the eliminativist. The most natural strategy seems to rely on the fact that, in the appropriate limit, QG degrees of freedom should approximate GR. Thus, we can, for specific notions such as localization, define approximate QG versions of localization which hold in such limits and justify using localized experimental setups.

In this chapter, I will test these ideas in the context of the AdS/CFT correspondence, the leading approach to defining a holographic theory of QG, paying particular attention to the description of the interior of evaporating black holes. In particular, I am going to argue that in this context, spacetime emergence as a whole does not work,[2] and provide a different account of the relation between QG and GR spacetime, which I call *operational recovery*, somewhat related to the eliminativist proposal of Baron (2021), though more extreme in spirit since, as I am going to argue, only operational data makes sense when it comes to the semiclassical regime of AdS/CFT where spacetime was supposed to emerge. I will also propose two metaphysical glosses on operational recovery, *operational eliminativism* and *operational emergence*, which differ mainly in whether one takes spacetime to be reducible to operational data, and show how AdS/CFT, understood via operational recovery, avoids empirical incoherence.[3] Moreover, by discussing in detail the extent to which spacetime can be said to be emergent in AdS/CFT and how to characterize this emergence, I will tackle an issue which has received surprisingly little attention in the literature thus far. In particular, I will show that the *bulk interior reconstruction program* provides some critical and hitherto unexplored case studies on the issue of the emergence of spacetime in QG.

However, let me first address a possible skepticism regarding this project: one might object that, since we know that our world is not AdS, there is no need to understand whether spacetime emerges in this theory since this case study

[2]Baron (2020) presents an argument against spacetime emergence which is complementary to the present one: while Baron (2020) argues against the metaphysical consistency of (certain kinds of) spacetime emergence, here I present an argument that a specific physical construction of spacetime defies any emergentist interpretation.

[3]I will not discuss the hard problem since operational recovery avoids the issue entirely since it is only committed to empirical data, not any ontologically thick characterization of spacetime (in this being close to eliminativism), and thus presumably does not suffer from such issues.

cannot tell us anything about our world. More generally, one might worry that AdS/CFT is ultimately just a speculative physics thesis, and thus we should be wary about trusting it. I take this remark to be mistaken for two reasons. First, as mentioned in chapter 1, there are holographic constructions describing cosmological spacetimes, which show features analogous to those I discuss here, such as quantum islands and entanglement wedge reconstruction (Hartman et al., 2020; Chen et al., 2021; Balasubramanian et al., 2021; Bousso and Wildenhain, 2022). I rely on AdS/CFT because, in this context, the various notions are much better understood and well-defined than those models. However, there is reason to believe that these constructions generalize to the cosmological case. Second, as I mentioned already throughout this thesis, I believe it is helpful to focus on AdS/CFT because it is one of the few known non-perturbative formulations of QG. From this point of view, it provides an invaluable case study to understand precisely how spacetime is treated in QG.

The chapter is structured as follows: §5.1 introduces bulk reconstruction and explains the role that state-dependent operators and complexity theory play in it; §5.2 discusses the impact that these notions, and state-dependence, in particular, have on regarding semiclassical spacetime as emergent; §5.3 considers how complexity theory still afford a solution to the problem of empirical incoherence even without an emergent spacetime while discussing the possible metaphysical interpretation of this schema; §5.4 then concludes.

5.1 Bulk Reconstruction and Spacetime

This section discusses how spatiotemporal quantities are recovered from broadly non-spatiotemporal data in AdS/CFT. I will start in §5.1.1 with an overview of the program of bulk reconstruction; I will then move in §5.1.2 to consider a model where the features of bulk reconstruction relevant to my discussion, i.e., state-dependence and complexity theory, can be seen concretely.

Recall, as I remarked in the previous chapters, AdS/CFT is the duality between QG defined in AdS spacetime, i.e., the maximally symmetric solution with a negative cosmological constant of the Einstein Equations, in $n+1$ dimensions, and the CFT in n dimensions defined on the asymptotic boundary of the AdS spacetime. Since the CFT is well-defined non-perturbatively, this duality provides a non-perturbative definition of QG in AdS spacetime.

5.1.1 Bulk Reconstruction

In the context of AdS/CFT, most of the philosophical literature has focused on the issue of whether the gravitational bulk spacetime is emergent from the boundary CFT. There is, however, an immediate issue also of spacetime emergence within such theories. As is customary in AdS/CFT, the boundary CFT allows the definition of the theory's non-perturbative dynamics. Indeed, for this chapter, we will equivalently speak of CFT or fundamental description, and AdS or semiclassical/effective field theory description. We do this because, for us, gravity in the bulk will always mean semiclassical gravity, and boundary CFT will always mean the Hilbert space and correlation functions which represent the non-perturbative description of QG, thus far only known through the dual CFT. However, this language does not imply any claim of relative fundamentality between the (yet unknown) fundamental theory of QG in AdS and its holographic dual CFT description. Rather, the only claim of relative fundamentality made here, and the way fundamental should be read in this chapter, is between the fundamental description, currently known only in terms of the boundary CFT (which is dual to, not more fundamental than, the unknown AdS QG theory) and the effective description given by AdS semiclassical gravity (not full QG).[4] Since the non-perturbative dynamics given by the fundamental description are defined in terms of the CFT path integral, they will at most depend on the conformal structure and not include any spacetime structure stronger than this. Moreover, if we take the common core approach to the duality[5] (arguably the most popular interpretation of the duality among philosophers) according to which our fundamental ontology should only involve features which are invariant among the CFT and AdS sides of the duality, then also the dimensionality of spacetime will turn out to be unphysical. It then seems that the fundamental structures of AdS/CFT do not admit any straightforward spacetime

[4]Intuitively, one can think of this fundamentality relation in the same way as the relation between a UV complete fundamental theory and its low energy effective field theory.

[5]See Butterfield (2018) for a detailed articulation of this position.

interpretation, since paradigmatic spatiotemporal features such as distances (due to the conformal symmetry of the fundamental degrees of freedom) and dimensionality do not make sense at the fundamental level. Hence, we need to identify an appropriate way for spacetime to emerge.[6]

The standard way to understand the recovery of spatiotemporal degrees of freedom, minimally understood as given by an effective field theory in the bulk reproducing semiclassical gravity and thus GR, is the program of *bulk interior reconstruction* (Harlow, 2017, 2018; Penington, 2020; Akers and Penington, 2022). Bulk reconstruction is a program aimed at understanding how local bulk quantities defined only at the level of semiclassical gravity are represented in the non-perturbative definition of QG encoded in the CFT. This task is crucial to understanding the emergence of spacetime in AdS/CFT since it connects the emergent quantities of the bulk semiclassical gravity theory, which represent, in this setup, spacetime with its geometry, with the fundamental, non-perturbative quantities defined in the CFT, which in this context are the non-spatiotemporal quantities from which spacetime should emerge.

The crucial concept for the bulk reconstruction program is *entanglement wedge reconstruction*, which we already briefly encountered in chapter 2 while

[6]Indeed, even if one were to be skeptical of the fundamental degrees of freedom of AdS/CFT being non-spatiotemporal, it is still the case that presumably an extension of holography to asymptotically flat and de Sitter spacetimes will involve non-spatiotemporal fundamental degrees of freedom, due for example to the boundary structure of those spacetimes (Witten, 2001). Since, moreover, bulk reconstruction is supposed to be a general feature of holographic theories (Bousso and Penington, 2022), the main thrust of this discussion would still hold for these more physically relevant cases.

discussing the resolution of the AMPS paradox for evaporating black holes developed in Penington (2020). Let me now delve deeper compared to chapter 2 into entanglement wedge reconstruction. In order to define the entanglement wedge of a given boundary region R, we first need to know what the quantum extremal surface **(QES)** associated with that region is. We define the **(QES)** as follows:

(QES) A **quantum extremal surface** χ is defined as a surface satisfying two conditions:

(i) *Homology Constraint:* given a boundary region B, a surface χ satisfies the homology constraint if, for C a space-like hypersurface, $\chi \cup B = \partial C$, i.e., the union of χ with a boundary region B is the boundary of some space-like region C. C is called *homology hypersurface*.

(ii) *Extremize generalized Entropy:* the surface χ should be a surface which extremizes the generalized entropy

$$S_{\text{gen}}(\chi) = \text{ext}\left[\frac{A(\chi)}{4G_N} + S_{\text{bulk}}(\chi)\right] , \tag{5.1}$$

where $S_{\text{bulk}}(\chi)$ is the von Neumann entropy[7] of the bulk fields contained in $\chi \cup B$ and $A(\chi)$ is the area of the hypersurface χ.

[7] If the total bulk state is pure, then this von Neumann entropy is the entanglement entropy between what is inside and what is outside the entanglement wedge.

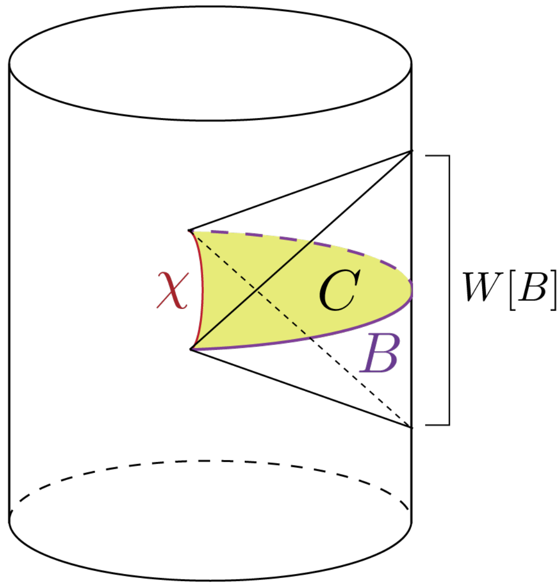

Figure 5.1: An example of entanglement wedge. Here $W[B]$ is the entanglement wedge of the boundary region B (purple), χ (red) is the **QES** which minimizes S_{gen}, and C (yellow) is the homology hypersurface.

We then define the quantum Ryu-Takayanagi or HRT surface as the QES associated with the boundary region R, which minimizes the generalized entropy. Given the notion of HRT surface, we can then define the notion of Entanglement Wedge **(EW)** (figure 5.1), as follows:

(EW) Let B be a boundary spatial subregion of an asymptotically-AdS spacetime. The **entanglement wedge** of B, which we denote by $W[B]$, is the bulk domain of dependence $D[C]$[8] of the homology hypersurface C

[8]The domain of dependence $D[C]$ is the set of points with the property that any causal

delimited by the HRT surface χ.

We then arrive at the crucial tool for bulk reconstruction, which is the entanglement wedge reconstruction conjecture, which says that

(EWR) Entanglement wedge reconstruction: all physical quantities in $W[B]$, i.e., the entanglement wedge of a spatial subregion B, are represented in the CFT by operators in B.

With the notion of entanglement wedge reconstruction, we now have a way to represent local, semiclassical bulk quantities in the fundamental language of the boundary CFT. In particular, the relation between $W[R]$ and R is given by a quantum error correcting code,[9] which is needed in order to encode the locality properties of the bulk degrees of freedom in their CFT description. While the details of this statement go beyond the scope of this chapter, what is essential to the present discussion is that the Hilbert space of the semiclassical bulk gravitational theory in $W[R]$ is mapped not to the entire Hilbert space of the boundary region R, but only to a specific subspace, which we call the code subspace. This code subspace is the space of bulk semiclassical quantities and represents semiclassical gravitational degrees of freedom in the CFT. The quantum error correcting code gives us a map between bulk semiclassical quantities and their representation in the CFT.

curve passing through one of these points must also intersect C.

[9]See Almheiri et al. (2015) for the derivation of the quantum error correcting code structure of the holographic map, due to issues having to do with locality in the bulk semiclassical approximation. See also Bain (2020) for philosophical discussion of these constructions.

There is a catch to this statement, however. As I have introduced things thus far, the map between the entanglement wedge $W[R]$ and R is treated as an exact equivalence, that is, as a linear isometry between Hilbert spaces. This equivalence does not, however, obtain in general: the map is only known in perturbation theory in G_N, Newton's constant. Therefore, we only know the CFT representation of bulk quantities as a perturbative sum. In particular, the effect of non-perturbative corrections, which we know are encoded in the CFT, is that the map between semiclassical and fundamental quantities is not even approximately an isometry. Equivalently, the inner product of semiclassical gravity counts as orthogonal states that are not orthogonal in the fundamental description, which thus has a Hilbert space of lower dimension compared to the expectation of semiclassical gravity. Usually, such non-perturbative corrections can be ignored as long as one deals with semiclassical quantities, as we are doing here, since they are small. Equivalently, we can treat the error-correcting map as an approximate isometry. However, not even approximate isometry is always the case for bulk reconstruction. There are situations in which these non-perturbative corrections have to be explicitly taken into account, and the result is that they make the map between bulk semiclassical gravity and the fundamental non-perturbative description that the CFT affords us, not isometric. Thus, they destroy the possibility of representing bulk quantities in the CFT. Let us see why.

The most relevant example of this phenomenon is the behavior of an evaporating black hole in AdS spacetime after the Page time and the related interior reconstruction. In particular, when dealing with evaporating black holes the entanglement wedge $W[\mathcal{H}_{rad}]$ associated with Hawking radiation involves so-called *quantum islands* after the Page time (Penington, 2020; Almheiri et al., 2019), see figure 5.2. The appearance of said islands means, via entanglement wedge reconstruction, that the interior can be reconstructed from Hawking radiation. This phenomenon is highly non-classical and manifests a way in which standard spacetime thinking gets modified in QG, already at the semiclassical level, due to the peculiar properties of the reconstruction map.

5.1.2 A Concrete Example

While this example is the most physically intriguing case, for our purposes, a much simpler model will be sufficient to describe the relevant features of bulk reconstruction (see Akers et al. (2022) for this model and further discussion of these constructions). Take a quantum system r represented by the Hilbert space $\mathcal{H}_r$. We can think of it as giving a finite-dimensional analogue of the semiclassical description of the interior partners of Hawking radiation, i.e., of the black hole interior, for a fixed Cauchy slice at time t.[10] Also, define a quantum system B with associated Hilbert space $\mathcal{H}_B$, which we take as giving the fundamental non-perturbative representation of the black hole.[11] The semiclassical and the fundamental description of the radiation in the

[10] Appropriate covariant and dynamical generalizations of this most simple example exist, see Akers et al. (2022).

[11] Which in full AdS/CFT is defined through the dual CFT.

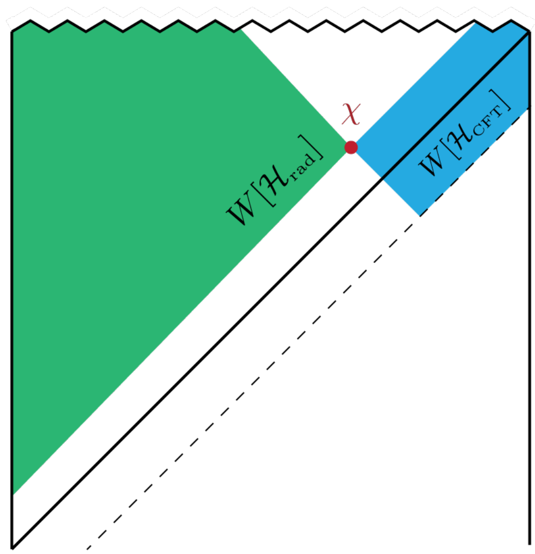

Figure 5.2: The Penrose diagram for a quantum island in an AdS black hole. The green region is the entanglement wedge $W[\mathcal{H}_{rad}]$ of the Hawking radiation, while the blue region is the entanglement wedge of the black hole. The HRT surface χ is in red. It is clear from the picture that (most of) the interior of the black hole is contained in the entanglement wedge of $W[\mathcal{H}_{rad}]$.

exterior of the black hole can also be included, in the form of a Hilbert space $\mathcal{H}_{rad}$ which is the same between semiclassical and fundamental description. However, taking into account the exterior radiation will not be necessary to introduce the main elements necessary to the ensuing discussion, so for the sake of simplicity I will omit mention of it. We define the map between semiclassical and fundamental description $V : \mathcal{H}_r \to \mathcal{H}_B$, implementing the reconstruction of bulk operators as:

$$V |n\rangle_r = \frac{1}{\sqrt{|B|}} \sum_b e^{i\theta(n,b)} |b\rangle_B \tag{5.2}$$

where $|n\rangle_r$ is a basis for $\mathcal{H}_r$, and $|b\rangle_B$ is a basis for $\mathcal{H}_B$, while $e^{i\theta(n,b)}$ are randomly chosen phases.[12] The map V tells us how the semiclassical description fits into the fundamental description. In particular, it has the structure of a quantum error correcting code, and its image in $\mathcal{H}_B$ gives the codespace encoding semiclassical information in the fundamental, non-perturbative Hilbert space.

Note that, for $|B|$ the dimension of $\mathcal{H}_B$ and $|r|$ the dimension of $\mathcal{H}_r$, the map V will be an isometry only for $|r| << |B|$. For $|B| << |r|$, there simply are not enough basis states in $\mathcal{H}_B$ to ensure that V be an isometry and thus preserve the inner product on $\mathcal{H}_r$. B has too small a dimensionality to allow V being an isometry. As we will see, the way that this failure of isometry manifests is by having small overlaps, in $\mathcal{H}_B$, between states which are orthogonal in $\mathcal{H}_r$. These non-zero overlaps imply that states with different spacetime geometries, thus counted as orthogonal by $\mathcal{H}_r$, are not fully distinct states in the fundamental description given by $\mathcal{H}_B$. In other words, the description given by spacetime geometry is not adequate to capture these small overlaps and hence fails.

Nonetheless, it should still be the case that the semiclassical description is

[12] The justification for the randomly chosen phases comes from the chaotic dynamics of the black hole, see Akers et al. (2022).

approximately valid. Indeed, Akers et al. (2022) show that:

$$\langle n'| V^\dagger V |n\rangle = \frac{1}{|B|}\sum_b e^{i\theta(n,b)-i\theta(n',b)} = \begin{cases} 1 & n = n' \\ O(1/\sqrt{|B|}) & n \neq n' \end{cases} \tag{5.3}$$

This equation means that when semiclassical basis states have non-zero overlaps in the fundamental description, this non-zero overlaps are of order (the inverse of the square root of) the black hole entropy. Since the black hole entropy is very large, this means that these overlaps will in general be very small.

Thus, even when $|B| << |r|$, it is still the case that, for basis states like $|n\rangle$, the semiclassical description works, i.e., the map V is an approximate isometry, up to order $O(1/\sqrt{|B|})$. Moreover, the QES formula and quantum islands can already be derived in this simple model (see Akers et al. (2022) for details on how the derivation would work). This line of thought leads to the idea that when working with a semiclassical approximation, we can have a large number of approximately orthogonal states, much more than there are in the fundamental description.

One might worry however that the preceding discussion is not enough to ensure the validity of the semiclassical description in all relevant regimes. Indeed, all (5.3) shows is that overlaps between semiclassically orthogonal states are small when such states are basis states. But certainly not all applications of semiclassical gravity can be analyzed in terms of basis states.

We do have superpositions also in semiclassical gravity, after all.

Indeed, something stronger is the case. Akers et al. (2022) show that the basic idea behind (5.3), i.e., that overlaps in the fundamental description between states orthogonal in the semiclassical description are small, is not limited to basis states and can be extended to any combination of such basis states obtained by applying an operator whose computational complexity is less than exponential, i.e., to states of *subexponential complexity.*[13] Hence, for states of *subexponential complexity*, the semiclassical description holds up to exponentially small corrections, even if $|B| << |r|$. In other words, for states of subexponential complexity, the difference between their overlaps in the semiclassical and fundamental descriptions is exponentially small in the black hole's Hilbert space dimension (or, equivalently, its entropy). In complexity theory, it is customary to associate exponentially complex operations with operations that can only be carried out in principle but not in practice, while subexponential ones can also be carried out in practice. This distinction stems from the fact that exponentially complex require an amount of resources to be carried that scales exponentially with the growth of a specified quantity,[14] and thus grows (prohibitively) fast, while resource consumption for subexponentially complex operations grows much more slowly. Therefore, this result seems to imply that failures of isometry go beyond any semiclassi-

[13]States prepared from a reference state by a quantum circuit of subexponential complexity, see the discussion below for the relevant definitions and Akers et al. (2022) for technical details.

[14]In computation theory, it is usually the input's length. We will see below how to apply this notion to black holes.

cal observer's concrete ability to detect them and thus do not threaten their use of semiclassical reasoning.

While detailed definitions and derivations of these claims can be found in Akers et al. (2022),[15] it is still helpful to qualitatively see how this would work. In what follows, a quantum circuit is a quantum analog of a classical circuit, i.e., a map mapping input qubits to output qubits through a specified series of operations. These operations include in particular quantum gates, i.e., unitary operators, usually on a few qubits, and projection operators.[16] If a state can be prepared by acting with a quantum circuit of subexponential complexity on a reference state, typically a type of vacuum state,[17] let it be a state of subexponential complexity. The complexity of the state here is quantified in terms of the number of quantum gates making up the circuit. We can then define the subspace of states in $\mathcal{H}_B$ which have subexponential (in the black hole's entropy) complexity, i.e., whose complexity scales as $log|B|$. For these states, a formula of the following kind holds:

$$Pr\left[sup_{|\psi\rangle,|\phi\rangle \text{sub-exp}} |\langle\psi| V^\dagger V |\phi\rangle - \langle\psi|\phi\rangle| > \sqrt{18}|B|^{-\gamma}\right] \lesssim \exp\Big(|B|^\alpha log\ log(|r|) - \frac{|B|^{1-2\gamma}}{24}\Big) \tag{5.4}$$

where γ and α are two indices whose precise definition is not relevant to the present discussion, Pr stands for probability, and all other terms have been defined above. The essential meaning of (5.4) is that, for an observer

[15] See also Brown et al. (2020); Engelhardt et al. (2021, 2022) for previous work on the connections between AdS/CFT, complexity, and black holes.

[16] See Nielsen and Chuang (2010) for a standard introduction.

[17] If we are working with n qubits, this state would be the $|0\rangle^{\otimes n}$ state.

only capable of performing operations of subexponential complexity, as we presumably are, the overlap between states corresponding to states in the semiclassical description, i.e., states in the codespace, will be exponentially small and thus not relevant to any concrete measurement they could make. In other words, the validity of the semiclassical picture, even for $|B| << |r|$, is protected by computational complexity.

To better understand this result, recall that the overlaps between semiclassically orthogonal basis states are of order the exponential of (the square root of) the black hole entropy, by (5.3). Hence, to be able to see these overlaps, one would need to take a linear combination involving an amount of basis states of order the exponential of the black hole entropy. Equivalently, one needs to act on a reference vacuum state (in the black hole case, this might be the Schwarzschild vacuum) with a number of quantum gates (unitary operators acting on a few qubits) of order the exponential of the black hole entropy. Given that each quantum gate acts on a few qubits, and that in the simple model we are looking at we are approximating the black hole with a qubit system whose fundamental degrees of freedom are given by $\mathcal{H}_B$, this is the same as saying that we need to act on the black hole degrees of freedom a number of times of order the exponential of the black hole entropy. Given the large entropy of a black hole, any such operation would be incredibly difficult to carry out in practice, since it would involve an enormous amount of interventions on the black hole.

Indeed, any operation, or more precisely any quantum circuit, i.e., a combination of quantum gates, able to construct a state which manifests the effects of these small overlaps (very roughly, such that $\langle\psi| V^\dagger V |\chi\rangle \neq \langle\psi|\chi\rangle$) will be a quantum circuit made of an amount of quantum gates of order the exponential of the black hole entropy. Hence, it will be a complex operation, in the sense that its complexity, determined by the number of quantum gates it involves, is exponential in the black hole entropy. Since it is customary in complexity theory to count complex operations as operations that cannot be carried out *in practice*, no real observer will be able to detect the non-zero overlaps, as doing this would involve an operation that cannot be carried out in practice, i.e., a complex one. On the other hand, as long as one is limited to operations of subexponential complexity, i.e., operations which can be carried out in practice, the overlaps will be invisible, and the semiclassical description valid.

Indeed, we can see the connection with measurements explicitly by thinking of a very simple model of measurement interactions. Suppose we wish to measure a certain quantum system ψ and that we have for this task a measuring apparatus χ. The composite system of quantum system and measuring apparatus starts in the state:

$$\sum_i C_i |i\rangle_\psi |0\rangle_\chi \tag{5.5}$$

where $|i\rangle_\psi$ is an eigenstate with eigenvalue i for some operator X acting on ψ, and $|0\rangle_\chi$ is the neutral, ready state for the measuring apparatus χ.

We implement the measurement by allowing the measuring apparatus χ to become correlated with the quantum system ψ. This evolution is represented by some unitary operator U_{meas}, and leads to the following state:

$$\sum_i C_i \left|i\right\rangle_\psi \left|x_i\right\rangle_\chi \tag{5.6}$$

where $\left|x_i\right\rangle_\chi$ is the state of the measuring apparatus χ corresponding to a measurement of i.[18] In the present context, if we take ψ to be a black hole, we see that (5.4) implies that, as long as the unitary U_{meas} implementing the measurement process is a quantum circuit of subexponential complexity, i.e., insofar as we limit ourselves to measurements that an observer can perform in practice, then the semiclassical picture will apply up to exponentially small errors. In formulas, and assuming for simplicity that ψ is in the eigenstate $\left|i\right\rangle_\psi$ for the operator X:

$$U_{meas}(V(\left|i\right\rangle_\psi) \left|0\right\rangle_\chi) \approx \left|i\right\rangle_\psi \left|x_i\right\rangle_\chi \tag{5.7}$$

where we implement both the subexponentially complex unitary U_{meas} and the holographic map V, and where $\approx$ holds up to errors of order $O(1/\sqrt{|B|})$, as in (5.4). Hence, in practice, an observer is not able to perform a measurement which detects the failure of semiclassical gravity.

[18] To be complete, this account should be amended to take into account of one's favorite solution to the measurement problem. Given however that the present discussion does not hinge on details of the resolution of the measurement problem, and moreover that the status of the measurement problem in QG is not known, I will omit discussion of the measurement problem for the purposes of this paper.

An important consequence of the map V not being an isometry is that bulk reconstruction will be in general *state-dependent* or *state-specific* (Papadodimas and Raju, 2013, 2016; Akers and Penington, 2021, 2022), a feature that we already encountered, though in a more model-dependent, less general form compared to the present setting, in chapter 4.[19] Recall from chapter 4 that by state-dependent, I mean that operators of the semiclassical description only have equivalents in the fundamental description by assuming a specific state. Formally, this means that for O an operator in the semiclassical description and $\widetilde{O}$ an operator in the fundamental description, it is the case that

$$\widetilde{O}V\,|\phi_1\rangle = VO\,|\phi_1\rangle \tag{5.8}$$

$$\langle\phi_2|\,V^\dagger\widetilde{O}V\,|\phi_1\rangle = \langle\phi_2|\,O\,|\phi_1\rangle\,, \tag{5.9}$$

which means that $\widetilde{O}$ gives the fundamental representation of O, hold only given the choice of a specific state $|\phi\rangle$. To make this dependence explicit, we will write $\widetilde{O}_{|\phi\rangle}$. For different $|\phi\rangle$s there will be different $\widetilde{O}_{|\phi\rangle}$ that satisfy (5.8). In other words, the semiclassical operators are not well-defined, in particular linear operators in the fundamental description. We say that O cannot be a linear operator in the fundamental description because for different $|\phi\rangle$s, we

[19]To be precise, state-dependent is usually associated with the proposals of Papadodimas and Raju (2013, 2016), while state-specific with those of Akers and Penington (2021, 2022). While there are some subtle technical differences between these approaches (Akers et al., 2022), these differences will not be relevant to the present discussion, so I will use the terms interchangeably.

have different $\widetilde{O}_{|\phi\rangle}$ in the fundamental description representing O, and these different $\widetilde{O}_{|\phi\rangle}$ cannot be combined linearly into a single operator.

This observation creates a problem, since it seems to follow that the operator O cannot exist, since fundamentally there is no operator like it,[20] but at the same time it seems reasonable that there should be a semiclassical description even in the black hole interior, when V is not an isometry. This is the tension that introducing state-dependent operators resolves. As I mentioned above, a state-dependent operator $\widetilde{O}_{|\phi\rangle}$ representing O is an operator in the fundamental description which behaves like O (has the same eigenvalues and gives the same scattering amplitudes, (5.8)) when applied to the state $|\phi\rangle$, but not when applied to other states. The same procedure can be applied, each time with a *different fundamental operator*, for any other state in the fundamental description Hilbert space $\mathcal{H}_B$. Thus, a collection of state dependent operators $\widetilde{O}_{|\phi\rangle}$ for all $|\phi\rangle$ in $\mathcal{H}_B$ gives a fundamental representation of the semiclassical operator O, but at the price of depending on the details of the fundamental representation, since each state dependent operators requires the specification of a fundamental state, and different fundamental states give rise to different incompatible state dependent operators. We thus get a way to represent O in the fundamental description and justify the expectation that there should be a semiclassical description even when $|B| << |r|$. However, this representation of O is not really an operator in $\mathcal{H}_B$, because

[20] Equivalently, O is not the semiclassical limit of any fundamental QG quantity. Hence, it is unclear how can it be an operator in semiclassical gravity, if it is not the semiclassical limit of a QG quantity.

V is not an isometry. Rather, it is a non-linear combination of operators in $\mathcal{H}_B$, each behaving like O for a specified state, but not for the rest of $\mathcal{H}_B$, and which cannot be combined linearly into something looking like O (since there cannot be a linear operator in $\mathcal{H}_B$ representing O).

The impossibility of combining different $\widetilde{O}_{|\phi\rangle}$s into a single linear operator, and more generally the need for state-dependent operators, can be justified as follows. Given an operator O in the semiclassical description, this operator's eigenstates will, as usual, form a basis for $\mathcal{H}_r$. However, if $|B| << |r|$, these vectors cannot all be basis vectors in the fundamental description given by $\mathcal{H}_B$. Thus, the operator O cannot straightforwardly exist in $\mathcal{H}_B$. However, we can resolve this issue if we allow the fundamental representation of O, $\widetilde{O}$, to be the state-dependent operator $\widetilde{O}_{|\phi\rangle}$. In this case, all we require is that (5.8) holds for a specific state $|\phi\rangle$, not for all states in the codespace. Thus, we have no problems with a mismatch in the dimensionality of r and B. For this reason, state-dependent operators can be well-defined even if $|B| << |r|$, while their state-independent counterparts cannot, for the reasons just given. Thus, no single linear operator represents O in the fundamental description; consequently, O cannot be a linear operator at the fundamental level. As a consequence, state dependent operators end up depending on the details of the fundamental representation, in the sense that their definition requires the specification of a fundamental state, and that different fundamental states give rise to different state dependent operators. Thus, semiclassical operators do not have a non-perturbative definition (which would require the existence of an operator in the fundamental description) but only make sense for (small perturbations around) a given state (the state in state-dependent), i.e., in a perturbative expansion.[21] In particular, this will be true for $|B| << |r|$ when non-perturbative effects become relevant, leading to V not being an isometry. For $|r| << |B|$, there is no reason to expect large-scale failures of isometry in V and thus similar drastic consequences for the semiclassical description.

Prima facie, it would seem that the appearance of state-dependent, non-linear operators, and thus the large-scale failure of the semiclassical approximation for the interior of black holes after the Page time, would be proof of the inconsistency of the whole holographic approach. Indeed, black holes have interiors![22] However, recall that we have seen in (5.4) that the semiclassical description is still approximately accurate, at least as long as we restrict ourselves to states with subexponential complexity, even when $|B| << |r|$. Thus, there is no contradiction between state-dependent operators and the semiclassical description of the black hole interior. Indeed, state-dependent operators are necessary to accommodate significant failures of isometry in V, while the exponential complexity of detecting such failures ensures that no observer will detect such failures, since this would require an operation that they cannot in practice carry out. Hence, they will be able to use the semiclassical approximation to describe the interior, even if, strictly speak-

[21] One can see this fact also from the discussion in chapter 4.

[22] At least insofar as one wants to avoid firewalls (Almheiri et al., 2013).

ing, its operators are not well-defined since such behavior goes beyond what they can observe.

The picture that emerges is that, in AdS/CFT, a non-isometric, approximate quantum error correcting code gives the relation between fundamental and semiclassical description. Such code is at times approximately isometric, and sometimes, in particular, inside black holes after the Page time, it becomes significantly non-isometric. Therefore, the semiclassical approximation fails in these situations, and this failure manifests in the appearance of state-dependent operators, which implies that semiclassical operators are not well-defined linear operators in the fundamental description. Such failures of isometry, however, are not detectable because doing so would entail performing a computationally complex operation, which goes beyond the capacities of any observer. Thus, observers are still allowed to use semiclassical reasoning, as long as they confine themselves to subexponential operations, despite such reasoning being strictly speaking ill-defined inside black holes.

5.2 Emergence and State Dependent Operators

In this section, I will evaluate the relation between the fundamental description and spacetime in the semiclassical description. In particular, I will suggest that emergence is not the correct relation to describe how the semiclassical and fundamental descriptions are related due to the presence of state-dependent operators. Note that for this chapter, I will identify spacetime with the spatiotemporal structure of semiclassical gravity, i.e., the effective field theory corresponding to the perturbative quantization of GR (Wallace, 2021). In AdS/CFT, we will deal with semiclassical gravity in AdS spacetime. This definition implies that, by spacetime, we will usually refer to some kind of n-dimensional Lorentzian manifold or some structure of that kind, as is customary in the literature on spacetime emergence.

The proposal that emergence is not the correct relation between GR ad QG (and relatedly that spacetime, in general, does not exist even non-fundamentally) should strike the reader as somewhat unexpected. Indeed, emergence has been touted as the standard way to relate general relativistic spacetime to QG's fundamental, non-spatiotemporal goings-on. Indeed, some, such as Lam and Wüthrich (2018, 2021b), have even suggested that a specific form of emergence, functionalism, is the only game in town to account for the relationship between QG and spacetime (though see Baron (2021); Baron and Le Bihan (2022a) for possible alternatives and Linnemann (2021) for skepticism that functionalism is the right relation). In particular, emergence has been regarded as the best hope to solve QG's so-called *empirical incoherence problem*. I will discuss this issue in the following section, while for the remainder of this section, I will argue that emergence is not a fitting relation for the black hole interior in AdS/CFT.

First, let me briefly describe what I will minimally take emergence to be for this chapter. Such a minimal characterization, while inevitably entailing that no specific proposal for emergence will be considered in its details, will nonetheless allow me to state my argument in the most general way possible. There are a variety of accounts of emergence in the philosophy of physics and metaphysics literature.[23] These accounts differ, for example, in what they take the emergence relation to be exactly, whether they take it to be compatible with reduction or not, or in whether they take emergence to be compatible with physicalism. All these accounts, however, tend to have in common the belief that emergence is a phenomenon encompassing two prima facie contrasting aspects:

> **Autonomy:** emergent entities are in some relevant sense autonomous from, or novel with respect to, the entities from which they emerge. In particular, whatever there is to be said about emergent entities goes beyond what can be said about the microscopic entities from which they emerge and ensures that said entities have a degree of autonomy from the microscopic entities, in the sense that, for example, causal powers can be attributed to them directly (Wilson, 2015)

> **Dependence:** emergent entities depend on the microscopic entities from which they emerge. In particular, the emergent entity is not wholly distinct from that from which it emerges but instead owes its properties, in some sense to be specified, to the properties of the entities from which it emerges. An example of the relation of ontological

[23]See O'Connor (2021) for a review.

dependence between emergent and microscopic entities, which has been argued to be relevant to emergent spacetime, could be grounding (Wilson, 2021). Another relation, which has been discussed in the context of approaches such as that of Butterfield (2011a,b), is reduction in the sense of there being a formal derivation of the emergent from its fundamental base.

The idea behind emergence is that there should be a relation capable of fulfilling these two roles simultaneously and that this relation is crucial to understanding the relations between entities, as described by scientific theories, at different levels of fundamentality.

A good example, though not free of complications,[24] to understand how emergence enters into physical theory is the relationship between thermodynamic and statistical mechanics, which I will now explain in very rough outline. Let us take a system of n particles, whose phase space $\hat{M}$[25] is then going to be $2n$ dimensional. Each point $p \in \hat{M}$ is a specific microscopic configuration of the n particles, i.e., a microstate, whose behavior is described via statistical mechanics. In this space, we can define thermodynamic quantities such as temperature or entropy as functions of phase space, which are constant on specific subregions of phase space, i.e., different microstates have the same entropy and temperature. We can think of this procedure as coarse-graining because by using entropy, temperature, and similar properties, we can only

[24] See Frigg (2007) for a discussion of the relevant issues.

[25] Formally given by the cotangent bundle T^*M of the manifold M where the n particles live.

identify a subspace of phase space, not a specific microstate. There is, thus, a limit to the resolution of phase space available through thermodynamic means.

We can immediately check how this coarse-graining, and thus the relationship between statistical mechanics and thermodynamics, can be understood in terms of emergence. First, let us look at **dependence**. Since each thermodynamic property is fixed by fixing the appropriate microstate, and since the microstates are presumably more fundamental than the coarse-grained thermodynamic properties, it seems natural to take the statistical mechanical description to be more fundamental than and to ground the thermodynamic description.[26] Moving to **autonomy**, we can similarly check that it holds thanks to the coarse-graining relation between microstates and thermodynamic properties. In particular, since a large variety of different microstates correspond to the same value for a thermodynamic property, we can say that the thermodynamic property is independent of the details of the microstates. In other words, the definition of thermodynamic properties is autonomous from the detailed structure of the statistical mechanical microstates and, at most, depends on their coarse-grained, large-scale properties. Thus, the coarse-graining between statistical mechanics and thermodynamics, in this very rough description we have given, is a straightforward instance of emergence, understood as the conjunction of **dependence** and **autonomy**.

[26] In Butterfield (2011a)'s approach, again one would have to amend this story to ensure that it extends to an appropriate formal derivation of thermodynamics from statistical mechanics.

With this example in mind, we can understand why state-dependent operators, and thus the semiclassical description in AdS/CFT, do not permit a description in terms of emergence. Let me start by listing the trouble that AdS/CFT creates for the **autonomy** requirement. Contrary to the thermodynamics case, the presence of state-dependent operators means that quantities in the semiclassical gravity description depend sensitively on the details of their fundamental description. In other words, when state-dependent operators are involved, we cannot ignore the fine-grained structure of the fundamental state of the system when trying to define semiclassical quantities. This dependence on the fine-grained structure of the fundamental quantum state comes from the fact, seen in §5.1, that the definition of a state-dependent operator involves a choice of a state in the fundamental description, and that for different states there will be different state-dependent operators. Hence, the structure of each specific fundamental state is crucial in the definition of each state-dependent operator and in ensuring that, when the system is in the appropriate state, the state-dependent operator behaves semiclassically. This fact is in sharp contrast with the thermodynamics case, where **autonomy** is guaranteed exactly because such sensitive dependence on the microstates is avoided since we associate each value of a thermodynamic quantity with a set of microstates, not with a single one. Thus, it seems that under no reasonable reading of the **autonomy** constraint, state-dependent operators can be construed as autonomous from their fundamental description.

When it comes to the **dependence** requirement, the situation is even more extreme. As we have seen in §5.1, state-dependent operators are not linear operators. Assuming (i) that a minimal condition for a relation of dependence is that both its relata exist and (ii) that for an operator to represent an actually existing, physically well-defined entity, it must be a linear operator,[27] then spacetime in the context of the black hole interior does not satisfy **dependence**. We have this conclusion because spacetime is, in the end, to be given in terms of state-dependent operators, and these operators are not linear. Since these operators, by the minimal requirement expressed above, cannot represent well-defined existing entities, they also cannot represent the relata of a dependence relation. In other words, spacetime does not exist, since state-dependent operators are not suited to represent something that exists. They are not suited to represent something that exists because they are non-linear, and quantum mechanics requires that an entity be represented via states in Hilbert spaces and linear operators. Since spacetime does not exist in the present context, it cannot be the relata of a dependence relation. Hence, **dependence** cannot be satisfied.

An analogous argument holds for approaches cashing out **dependence** in terms of formal derivation. Since formally well-defined operators in quantum mechanics are always linear, a non-linear operator is not a formally well-defined object in quantum mechanics. Hence, any formal derivation crucially employing non-linear operators, as in the present case, cannot sat-

[27] As is customary in quantum mechanics.

isfy **dependence** since what it is deriving is not the quantum mechanical operators it should have derived.[28] Thus, the **dependence** requirement cannot be satisfied.[29]

In conclusion, emergence is not, in general, the correct relation to describe the connection between semiclassical and fundamental descriptions in AdS/CFT, especially as it regards spacetime.

5.2.1 Objections

Before moving on, let me briefly discuss two ways one might try to avoid this conclusion and why they are not convincing. The first argument starts from the observation that in many discussions of state-dependent operators, state-dependent operators are shown to exist not only for a single fundamental state but for various fundamental states (see, for example, the proposals in Papadodimas and Raju (2013); Hayden and Penington (2019); Penington (2020), whose basic features were outlined in chapter 4). Wouldn't this fact be enough to establish **autonomy** in analogy with thermodynamics? The crucial issue, in this case, is that, while such subspace-dependent operators certainly make sense, the relevant subspace never covers the entire codespace,

[28]It should have derived a quantum mechanical operator because both the fundamental theory of QG and semiclassical gravity are quantum mechanical theories.

[29]This argument also shows that it is not enough to avoid the issues raised here to eschew emergence and move to a purely fundamentality-based, for example, ground theoretic, understanding of the relation between spacetime and QG (as arguably hinted at in Wüthrich (2017)). The reason is that the failure of **dependence** means that there is no appropriate fundamentality relation involving spacetime when it comes to the black hole interior in AdS/CFT.

or equivalently it never covers all states with a semiclassical interpretation. Thus, even if we could move to subspace-dependent operators, this move still would not remove a significant degree of dependence on the microstructure encoded in the fundamental description. While this proposal might ameliorate somewhat the strangeness associated with state-dependent operators, it cannot do so to a degree sufficient to establish **autonomy** in a sense relevant for emergence. Indeed, no subspace will ever do, since state-dependent operators appear due to $|B| << |r|$. Moreover, it is unclear how this proposal would deal with the failure of **dependence**. Thus, it is simply a matter of dimensionality of the Hilbert spaces in the semiclassical and fundamental descriptions that makes a state-dependent operator unsuitable to emergence. Another argument goes as follows. While state-dependent operators are hostile to emergence, it seems that they only appear in the fundamental description, while in the semiclassical description, they are simply operators in the semiclassical Hilbert space $\mathcal{H}_r$. Thus, or so the thought goes, state dependence should be seen as an artifact of the fundamental description and not impact our judgment regarding the emergence of spacetime and other semiclassical, and not fundamental, structures. The issue with this argument is that it assumes that we can discuss emergence not by thinking of the relation between fundamental and semiclassical descriptions, since doing so leads us back to state-dependent operators. Instead, we are invited to think of the semiclassical description as its own entity. This move, however, is untenable since emergence is a relation between fundamental and emergent descriptions. Emergence thus requires a specification of how the two are related to be discussed. Again, the comparison with thermodynamics is illustrative. Even if we could define thermodynamics quantities in a way fully independent of any statistical mechanical underpinnings, for example, via an abstract theory of entropy and temperature, this theory would not be enough to discuss emergence. Instead, we would still need to embed this theory into the statistical mechanical phase space and then evaluate the relationship between thermodynamics and statistical mechanics in that phase space. In the context of AdS/CFT, doing so means studying the holographic map V and, in particular, coming to terms with the fact that state-dependent operators represent semiclassical quantities. Thus, this argument cannot work to show that state-dependent operators would be ignored when discussing the emergence of the semiclassical description from the fundamental one.

5.3 Complexity and Empirical Incoherence

We have seen in the previous section that emergence is not the appropriate relation to connect the fundamental and semiclassical descriptions in AdS/CFT. Consequently, we cannot say that spacetime is emergent in AdS/CFT, at least not in general. The semiclassical spacetime in the interior of black holes does not emerge. This fact raises, however, a fundamental problem for the consistency of AdS/CFT itself. As I mentioned before, the main reason behind the idea of emergent spacetime is the need to answer the challenge of empirical incoherence (Huggett and Wüthrich, 2013). Empirical incoherence refers to the fact that a scientific theory might undermine, by its very structure, the very same predictions on which the theory relies for confirmation. A good example is provided by certain kinds of no-collapse interpretation of quantum mechanics, the so-called *bare theory* (Barrett, 1996, 1999). The idea would be that if we came to believe the bare theory based on certain experimental results, then the bare theory itself would tell us that we should not trust those experimental outcomes.

Interestingly, a similar problem has emerged in theories of QG, where the fundamental degrees of freedom are not spatiotemporal (Huggett and Wüthrich, 2013). In QG, the issue stems from the observation that spatiotemporal localization seems necessary to account for experimental data. In other words, it is unclear what an experiment would even be, let alone how to perform it, in the absence of spacetime structure allowing localization in space and time. Thus, it seems that any experimental evidence we could gather for QG would undermine itself, since QG itself tells us that such experimental evidence is not possible. The answer to this impasse put forward in Huggett and Wüthrich (2013) is that we should treat spacetime as emergent. If spacetime is emergent and thus available to entities like us, we can account for experimental data through this emergent spacetime, despite spacetime not existing at the fundamental level. Thus, QG avoids the problem of empirical incoherence by positing emergent spacetime, or at least so Huggett and Wüthrich (2013) contend.

It is now immediate to see what the issue is if, in AdS/CFT, emergence of spacetime is not available: how can we solve the problem of empirical incoherence in AdS/CFT? More specifically, how can we understand experiments and measurements concerning the black hole interior if the interior spacetime is not emergent? Indeed, if there is no emergent spacetime, then it seems that there are no experiments insulated from the non-spatiotemporal fundamental reality, and thus that no experiment as we are used to defining them is possible. I want to suggest that this is not a problem in the present context of AdS/CFT. We can see the reason, thinking back to the discussion in §5.1. The issue behind empirical incoherence is the meaningfulness of experimental procedures in a world without spacetime. However, the basic point behind (5.4) is that measurements and experiments carried out by, broadly speaking, semiclassical observers are indistinguishable from experiments carried out in semiclassical gravity. A fact which is true independently of whether spacetime is actually emergent or simply absent. Indeed, any localization or other spatiotemporal property that can be defined and influence the experimental procedures of an observer will be correctly recovered from the fundamental description via the holographic map V and state-dependent operators, up to exponentially complex operations. In particular, as long as our hypothetical observer is, as they presumably are, limited to operations of subexponential complexity, then (5.4) ensures that their measurements and anything they can learn about the world around them will look perfectly spatiotemporal, despite it not being so.

To see this, consider for example what localization in time concretely means. If I want to say that an experiment takes place at a certain place at a certain time, what I am really doing is measuring the values on a clock and a system of rods and correlating them with my experiment. As such, localization in spacetime is a matter of ensuring that a certain measurement has a certain result, which means, in a quantum theory, that a certain operator has an appropriate set of eigenvalues. Equation (5.8) ensures that if I try to carry out such a measurement, there will be an appropriate (state-dependent) operator whose eigenvalues will give my system of rods and clock readings, i.e., whose operational data is equivalent to that of a spatiotemporal theory. Moreover, complexity theory ensures that I am unable to see deviations, in my experimental data, from such spatiotemporal behavior. Nonetheless, even in this simple example, spacetime (and hence the place and time that my rods and my clock are supposed to be measuring, i.e., individuate) is not there, since state-dependent operators do not represent existing entities, as seen in 5.2 when discussing dependence. Hence, we have a time and place reading on our instruments (the clock and the system of rods) despite spacetime not really being there.[30]

As such, there is no problem of empirical incoherence even if, strictly speaking, there is no emergent spacetime, and thus presumably no spacetime, since we are still assured that all our measurements will look spatiotemporal, and

[30] Note that this is not supposed a full analysis of localization in operational terms. Rather, it is a reasonably intuitive example to understand how having access to only operational data can still be sufficient to understand notions such as localization.

thus data recovered from them will be perfectly acceptable. In other words, as long as the relevant observable quantities have the required properties, it is not a problem whether these observable quantities can sensibly be understood as spatiotemporal. We can summarize this point in the following criterion of *operational recovery*:

> **Operational Recovery:** as long as experimental information can be defined appropriately, chiefly in terms of the recovery of observables approximating the behavior of semiclassical gravity, then there is no problem of empirical incoherence, independently of whether spacetime is emergent.

Let me briefly sketch how to understand the metaphysics underlying operational recovery. I think that there are two ways to think about operational recovery in metaphysical terms, depending on whether one thinks that the operational data underlying the principle is sufficient to exhaust what there is to be said about spacetime or not:

> **Operational Eliminativism:** eliminativism about spacetime has been discussed in Baron (2021) and amounts to the thesis that spacetime strictly speaking does not exist, but rather QG gives an approximation to it in the appropriate regime. In the present context, operational eliminativism claims that there is no such thing as spacetime, but rather QG degrees of freedom, in the appropriate regime, give an arbitrarily good approximation to the observable data that would be typical of a

spacetime theory, thus avoiding empirical incoherence.

> **Operational Emergence:** by operational emergence, I mean the idea that insofar as spacetime is reduced to a given pattern of experimental data, one can recover a notion of emergent spacetime. As complexity ensures that we cannot see the effects due to the non-isometric nature of the holographic map V, we can rest assured that the observational data does not depend on them, thus ensuring that we avoid empirical incoherence.[31]

Let me briefly discuss the relation of these ideas with spacetime functionalism and its application to spacetime emergence in QG (Lam and Wüthrich, 2018). The main difference between operational recovery and spacetime functionalism, in this context, is that operational recovery is only committed to operational data, and hence does not assume the existence of spacetime. Proponents of spacetime functionalism in the context of QG, on the other hand, typically take their view to be a way to articulate the relation between spacetime and QG, and hence assume that there must be such a thing as spacetime, beyond the mere operational data associated with a spacetime theory, to which one is committed. While such a notion of spacetime is analyzed in functionalist terms, it is still the case that this notion of spacetime is a further posit compared to the commitments of operational recovery. A somewhat special case is given by operational emergence. In particular, in

[31]Indeed, it better be, for otherwise, semiclassical observers would not exist since they presumably do not have access to the fundamental picture.

the case of operational emergence one might suspect that there is a tighter relation with functionalism, in the sense that the reduction of spacetime to operational data would proceed in functionalist terms. While this is certainly a possibility, it is important to observe that the resulting sort of functionalism would be the different from the sort of spacetime functionalism usually advocated. In particular, spacetime would still not exist except for its operational content, and any further spatiotemporal notion would have to be expressed in operational terms. This approach would be in stark contrast with the more ontologically expansive approach advocated for example by Lam and Wüthrich (2018). While for them, it makes sense to speak of spacetime areas and volumes, and to analyze them functionally in terms of appropriate QG quantities, in the context of operational emergence one would first need to reduce (possibly through a functionalist analysis) spacetime quantities, such as areas and volumes, to operational data. That operational data would then be related to QG. While subtle, this detour through operational data is not optional, and is crucial to why operational emergence (and operational recovery more broadly) is applicable to cases such as the AdS/CFT one discussed here, while spacetime functionalism as usually understood, being a form of spacetime emergence, does not.

Besides issues with functionalism, one might worry that operational emergence, and more in general the discussion in this chapter, is in contrast with the discussion in the previous chapter regarding the recovery of semiclassical physics from amplitude data in state-dependent models of AdS/CFT. This is not the case, however, for two reasons. First, if instead of GR we used semiclassical gravity to define the spacetime causal structure, as would have arguably been more natural from the perspective of semiclassical physics and its relation to QG,[32] then the most natural relation to order the causal structure would have been a relation of *having nonzero amplitude in semiclassical gravity.* Given that semiclassical gravity amplitudes make sense in operational recovery as shown by (5.8), then there is no contrast between the claims of this chapter and those of chapter 4. Second of all, what the present discussion really shows is that the causal structures of spacetime and entanglement emerge from R_Q in a manner controlled by operational recovery, in the sense that they make sense only in the observation-dependent manner encoded in operational recovery. Hence, there is no issue in making operational recovery and the amplitude Humeanism of chapter 4 compatible.

Let me also note that the idea of operational emergence is not in contrast with the previous statement that spacetime emergence is untenable in general in AdS/CFT. When speaking of spacetime emergence, one usually has in mind that what emerges is something like a 4-dimensional Lorentzian spacetime or some structure like that. However, nothing like that is suggested here. Rather, spacetime is first reduced to a certain pattern of observable data, and only then we speak of emergence.

More interesting is the status of **autonomy** in this proposal. Prima facie,

[32] I avoided using semiclassical gravity in chapter 4 mostly to simplify the discussion, as using GR makes the connection to Humeanism easier to express.

one might worry that operational emergence is still prone to failure of **autonomy** due to state-dependent operators. Let me argue why this is not the case.[33] What I have argued thus far is that there is no well-defined entity corresponding to spacetime that we can say emerges insofar as we understand spacetime as represented by the operators and spacetime structure of semiclassical gravity. Nonetheless, I have also suggested, starting from the technical work of Akers et al. (2022), that the empirical data available in the interior of a black hole is equivalent to that available were semiclassical spacetime present. In particular, **autonomy** is guaranteed in this context by the fact that, up to exponentially complex operations, a description in terms of semiclassical gravity will be acceptable. Thus, as long as we limit ourselves to such operations, there is no need to bring the fundamental description into the picture. It is important, however, to remember that this procedure can only work if we restrict ourselves to the operational data available in the semiclassical picture. This is because the semiclassical picture's validity is given in terms of complexity theory, which encodes limits on actually implementable observation, not on the theoretical/ontological structure underlying such observations. Thus, in particular, we would be mistaken in saying that there is (emergent spacetime) in the sense of a 4-dimensional manifold since such a description is not well-defined due to state-dependence

[33] A parallel argument can be given for the satisfaction of **dependence**, based on the fact that the relevant observational data (and only it) is well-defined and depends on the fundamental QG structure. Given that the two arguments would proceed roughly the same, I will only go through them for the case of **autonomy**.

and makes sense only thanks to the operational restriction to subexponential operations in contrast with its supposed ontologically thick status. However, we can speak of (emergent) spacetime in the sense of a certain pattern of observational data obtaining, since such observational data is well-defined, thanks to complexity theory and, again thanks to complexity theory, does not depend sensitively on the fundamental description since we can use the semiclassical description to define it.

In conclusion, it seems that spacetime emergence can only be realized as operational emergence in the context of the black hole interior in AdS/CFT. Which of operational emergence and eliminativism one will choose then ultimately comes down to one's opinion on the feasibility of reducing spacetime to purely operational data. However, a discussion of whether this reduction is possible goes beyond the scope of the present chapter and will be left for future work.

5.4 Conclusions

We have seen in this chapter that the emergence of spacetime in AdS/CFT raises a host of profound and critical conceptual questions. In particular, it seems that the paradigm of spacetime emergence fails in this case and that the only sense that can be made of the appearance of spacetime is in purely operational terms. The extent to which this operational understanding is tenable calls for urgent philosophical attention. Indeed, while I have argued that empirical incoherence is avoided via operational recovery, it would be satisfactory to have a complete story of how spacetime quantities can be analyzed in operational terms, especially in light of proposals like operational emergence.

Another issue that immediately emerges once operational emergence is taken into account, an issue that is much closer to my overall conceit throughout this chapter with the status of Humeanism in QG, is the status of non-fundamental laws of nature in this picture, and in particular of the lawful content of GR. I will address this issue in the next chapter.

Chapter 6

General Relativity as a Special Science

In the previous two chapters, we have seen how AdS/CFT allows us to formulate Humeanism in QG, and how to recover spacetime, or at least its operational content, from these models, a crucial step in understanding how everything can supervene in Humean fashion on the fundamental QG degree of freedom. One thing that I have not discussed, however, is the impact of these discussions on the metaphysics of laws, one of the main topics within Humeanism, and in particular what is the status of the laws of GR vis-à-vis operational emergence and the other concepts discussed in the previous chapter. I will do this in this chapter.

Indeed, the metaphysics of laws of nature has proven to be fertile ground for the development of deep and detailed naturalistic metaphysics, and for testing the import of modern scientific theories on standard metaphysical theses. A particularly recent and interesting development on this front has been the attempt to evaluate the consequences of QG theories, and in particular of their supposed implication that spacetime fundamentally does not exist, on the metaphysics of laws of nature.[1] These developments took place, for the most part, within the broad framework of *naturalized metaphysics* (Ladyman et al., 2007) according to which metaphysical speculation should be based on the insights provided by our best science. As I have done throughout this entire thesis, I too will adopt the framework of naturalized metaphysics. Thus far, however, the literature has mostly looked into the status of fundamental laws in a non-spatiotemporal context, a topic which I have already addressed in chapter 4. Much less attention has been directed to the converse question of the status of special science laws in the QG context.

The goal of this chapter is to remedy this lacuna. A lacuna that is particularly egregious once we reflect on two facts. First, the status of special science laws is, if possible, even more controversial than the status of fundamental laws, to the point that some have even ventured to claim that special science laws do not count as genuine laws (Woodward, 2000; Cartwright et al., 2005); and even those, such as Cohen and Callender (2009), who accept that special science laws are indeed laws, have to then modify in highly non-trivial ways their underlying metaphysics of laws to accommodate for special science laws.

[1]See Lam and Wüthrich (2021a) for a detailed discussion of various views of the metaphysics of laws in QG, and Wüthrich (2020); Jaksland (2021); Matarese (2019) for discussion of the status of Humean approaches in QG.

Second of all, most of the scientific enterprise is indeed an enterprise in the special sciences. From condensed matter physics to biology to economics, the development of science is dominated by the development of the special sciences. This fact becomes especially relevant in QG, where the notion of special science ends up being applied also to GR and the QFTs of the standard model of particle physics.[2] Hence, in QG, if special science laws do not count as laws proper, then we face the somewhat bizarre predicament of having to say that the Einstein Field Equations are not laws of nature, contrary to what would presumably be basic physical intuition.

I will then in this chapter commence the study of special science laws in QG, by focusing in particular on the status of GR in QG, and argue, in this context, that broadly Humean theories of laws seem to be favored compared to their more modally committed counterparts.[3] This starting point is especially natural since one of the main conceptual hurdles in QG is making sense of the notion of spacetime as emergent. Insofar as the program of studying spacetime is to be successful, it should presumably include an explanation of the lawful (or unlawful) status of GR in QG. Moreover, the various notions of emergence thus far discussed in the philosophy of QG literature provide a

[2]For ease of exposition, throughout this chapter I will speak of specific theories, especially GR and QG, as if they were themselves sciences. Strictly speaking, QG stands for, in this context, fundamental physics, while the special science including QFT, GR, and semiclassical gravity, which constitutes the most direct descendant of QG physics, should be called something like *non-QG physics*.

[3]Berenstain and Ladyman (2012) argue that Humean approaches cannot account for special science laws. Insofar as the argument of this chapter is successful, it seems that the opposite is true and that at least sometimes the weaker modal commitments of Humeanism are necessary to make sense of the notion of special science laws.

natural starting point to evaluate the status of GR laws in QG.

In particular, in this chapter I will focus on *operational recovery*, a particular way of making sense of the relationship between spacetime and QG degrees of freedom according to which GR spacetime only makes sense operationally, that is, only the empirical predictions of GR are preserved by QG. Chapter 5 introduced operational recovery and argued that an approach of this type is required to make sense of recent work on the black hole interior in AdS/CFT (Papadodimas and Raju, 2013; Penington, 2020; Almheiri et al., 2019; Akers et al., 2022), where the holographic map connecting the fundamental QG degrees of freedom with the effective field theory of gravity leads to a breakdown of a straightforward notion of semiclassical geometry. As we are going to see in the following, operational recovery forces us to reconcile two contrasting impulses, a broadly operational understanding of spacetime and a realistic attitude towards the laws of GR. Such a combination, which to my knowledge has not been explored thoroughly in the context of the metaphysics of laws, naturally leads, as we are going to see, to a Humean approach to laws of nature.

More in detail, I will argue that the requirements of operational recovery lead to the surprising conclusion that only broadly Humean approaches to laws of nature allow us to make sense of GR as involving laws of nature. In particular, I will illustrate this general moral by showing that a specific Humean approach to laws of nature, the Better Best Systems (BBS) account of Cohen and Callender (2009), is uniquely suited to explain lawfully the relationship between spacetime and QG as conceptualized by operational recovery.

Hence, insofar as treating GR as involving genuine laws of nature is a desirable outcome, then Humean approaches are favored in QG, at least when the relevant physics of QG is close enough to a holographic theory.[4] Of course, this assumption is defeasible, and those not sympathetic regarding the prospects of holography or Humeanism will presumably want to push back on this point. On this issue, to which I will return more extensively in §6.3.3, let me for now simply mention that the use of holography in this chapter is more as an illustrative example of operational recovery, as testimony that it is a viable approach to understanding the relationship between spacetime and QG, connected with actual approaches to QG. Whether holography is ultimately the correct theory of QG does not in principle decide the truth of operational recovery as a way to conceive of the relation between QG and spacetime. Finally, even if operational recovery does not count ultimately as the correct way to understand the relationship between spacetime and QG, it still is a viable exercise at the present moment, where we still do not know what the correct theory of QG is, or what the correct way to think of the relationship between spacetime and QG is, to explore all possibilities. In this vein, I will explore what operational recovery can tell us about GR as a special science in QG.

The chapter is structured as follows: in §6.1, I recall the definition of operational recovery and its relation to holographic theories of gravity; in §6.2,

[4] Or instantiates operational recovery differently from the one displayed by AdS/CFT.

I argue that approaches to laws involving nomological necessity are in principle incompatible with counting GR laws as proper special science laws in this context; in §6.3 I consider and respond to some objections to this argument; in §6.4, I argue that Humean approaches fare much better, and do indeed succeed in counting GR laws as special science laws in the context of operational recovery; §6.5 concludes.

6.1 Operational Recovery and Holography

The goal of this section is to recall the basic features of operational recovery, as introduced in chapter 5, and its relation to AdS/CFT.

Recall from the previous chapter that we can understand operational recovery as follows:

> **Operational Recovery:** as long as experimental information can be defined appropriately, chiefly in terms of the recovery of observables approximating the behavior of GR (or semiclassical gravity), then GR is recovered from QG, independently of whether spacetime exists as an emergent entity.[5]

At its core, what operational recovery is saying is that, insofar as the experimental data typical of GR is recovered from QG, then in principle nothing else is required to understand the relationship between GR and QG. In the

[5]In chpater 5 I further distinguished between two metaphysical glosses on operational recovery, **operational eliminativism** and **operational emergence**. This finer distinction will not be relevant to the present discussion.

AdS/CFT context described in chapter 5, which will be the main theoretical context of interest to us in this chapter, it is immediate to see why operational recovery emerges naturally.

Semiclassical gravity, and hence spacetime, are only defined up to exponentially complex operations, i.e., up to the practical experimental capabilities of a realistic observer. Hence, we cannot appeal to ontologically loaded notions of spacetime emergence, and in particular, we cannot appeal to spacetime as a full-blooded emergent entity, to explain the relationship between GR and QG since spacetime in this case only exists insofar as observers cannot probe too finely. Hence, spacetime strictly speaking does not exist, since there are observables which do not have a spatiotemporal interpretation. However, realistic observers cannot access these observables, due to practical limitations encoded by computational complexity arguments. Hence, the appearance of spacetime is a function of the concrete, practical operational capabilities of observers, which presumably implies that we should not admit spacetime in our ontology.[6]

The observation behind operational recovery is that, in such a situation, experimental data is enough. Since such experimental data will explain all observations of our realistic, subexponentially complex observer, then, for such an observer, the world will indeed look like a world described by a spa-

[6] In principle, one also needs to take into account state dependence to fully appreciate this point, as remarked in the previous chapter. For the purposes of this chapter, however, it is sufficient to consider the role of complexity theory, which makes the exposition simpler. For this reason, I will bracket issues of state dependence and only focus on complexity and its relation to spacetime geometry.

tiotemporal theory like semiclassical gravity. That the world is not *really* like that is no concern for them, at least insofar as they are worried about the disappearance of spacetime wreaking havoc on their ability to carry out experiments and thus test QG, as in discussions of empirical incoherence (Huggett and Wüthrich, 2013). In other words, the operational recovery of spacetime gives back spacetime enough for all practical purposes and thus is sufficient to avoid troubles related to the disappearance of spacetime in QG. Having said this, let me move in the next section to a discussion of the metaphysics of laws and their relation to operational recovery.

6.2 Laws, Necessity, and Operational Recovery

In this section, I am going to discuss why theories of laws of nature that treat them as involving some kind of necessity or primitive modality are incompatible with operational recovery.

Let me start, however, by distinguishing what I will take, for this chapter, to be the core distinguishing feature of approaches to laws of nature involving necessity from those broadly Humean approaches which eschew such nomological necessity. Approaches involving nomological necessity are usually said to include theories of laws, such as:

- Approaches according to which laws of nature are necessary connections between universals, such as the approach of Armstrong (1983).

- Theories involving primitive powers and dispositions (Shoemaker, 1980; Bird, 2007). According to these approaches, laws of nature are a byproduct of the dispositional properties instantiated by various scientific entities, where by disposition I mean a property such as the fact that if a glass were to fall to the ground, it would shatter. Since dispositions are inherently modal, the necessity involved in dispositional properties is inherited by the laws that describe them.

- Theories treating laws of nature as primitive, sui generis entities, as in the approach championed in Maudlin (2007). Here, laws are treated as a new kind of entity, whose main job is to govern the evolution of concrete entities. Also, in this case, this notion of governing is taken to have a modal character, as stating something that *must* happen, not simply something that *does* happen.

What all these approaches have in common for present purposes is the idea that laws of nature encode or involve necessary connections between distinct entities. For example, when stating Coulomb's law that opposite charges attract and like charges repel, these approaches to laws wish to claim, in different forms and to varying degrees, that there is some kind of necessary connection holding between like and opposite charges. It is this necessary connection that underpins the specific behavior of specific instances of like-charged and oppositely-charged entities.

As such, we can take as a defining feature of approaches to laws involving necessity a violation of Hume's dictum that **(HD)** "there are no necessary connections between distinct entities".[7] Hence, for these kinds of approaches, **(NN)** *there are necessary connections between distinct entities.*[8] For this chapter, this will be the necessary condition for a theory of laws to count as postulating nomological necessity. In the same vein, the satisfaction of Hume's dictum is, for this chapter, the defining feature of approaches eschewing nomological necessity, broadly called Humean, that we will discuss in the next section.

Note that the approaches mentioned above, and more in general approaches to laws of nature involving nomological necessity, do not have to allow the treatment of special science laws as laws. In particular, it is an open question whether or not these approaches can deal with all features of special science laws, such as their being of limited scope or their admitting exceptions. While some approaches mentioned before do have extensions that cover special science laws, such as the dispositions-based account of Cartwright (1989),[9] some

[7]See Wilson (2010) for a detailed discussion of various formulations of Hume's dictum and its motivation.

[8]Note that, in principle, necessary connections here could be cashed out in different ways. For example, as placing a restriction on which possible worlds exist (only those with certain laws of nature), or as placing a restriction on which properties can be instantiated at which worlds (worlds where like charges repel instantiate the property of charge, while worlds where like charges attract instantiate the property of *scharge*). See Schaffer (2005) for an illuminating discussion of the various possible meanings of necessity here. For this chapter, I will remain non-committal on this point, as none of the arguments discussed here depends on such an assumption.

[9]Strictly speaking, Cartwright (1989) does not count special science regularities as laws. Nonetheless, insofar as she is accounting for special science regularities by positing some necessary connection in the world, her account falls into the scope of the present

are formulated exclusively for fundamental laws, such as Maudlin (2007)'s primitivist approach.

For this chapter, I will simply assume that there is an appropriate extension of theories of laws involving nomological necessity to special science laws. In case such an extension does not exist, given that instead such extensions exist for Humean theories (Cohen and Callender, 2009; Callender and Cohen, 2010), then this would preemptively close the discussion in favor of the Humean position.[10] Hence, for the sake of argument, I will assume that there exists a satisfactory account of special science laws involving nomological necessity.

Let me now argue why such a substantive understanding of special science laws is incompatible with operational recovery, and thus at the very least with the AdS/CFT cases where operational recovery naturally appears.

First, note that two conditions must be satisfied for **(NN)** to be true: (i) it must be the case that there are two (or more) distinct entities, and then (ii) there must be a necessary connection between them. While most theories of laws involving nomological necessity focus on characterizing the neces-

---

discussion and is mentioned for this reason. In particular, special science regularities still indeed absolve most of the functions expected of laws of nature, as remarked in Callender and Cohen (2010).

[10]At least insofar as one places a premium on counting special science laws as laws. While this assumption seems prima facie reasonable given standard scientific practice, as mentioned before some have called it into question (Woodward, 2000; Cartwright et al., 2005). In this chapter, I will assume that such a premium does indeed exist, given that it seems in keeping with standard scientific practice to count special sciences as involving laws. Given that, all things equal, it is better when working in a naturalistic framework to stick closer to scientific practice as possible, it is better to count special science laws as laws.

sary connection part, the existence of distinct entities should not be ignored. In particular, it is this component that fails when dealing with operational recovery. The reason is straightforward. If we are talking about GR (or semiclassical gravity) the two distinct entities among which a nomologically necessary connection should hold will be something like spacetime points.[11] However, as I have stressed before, when dealing with operational recovery there is no such thing as spacetime points available to us, if not as constructs out of operational, observer-dependent data.[12] Given the reasonable assumption that such operational constructions cannot deliver the kind of "distinct existing entities" that are required to make **(NN)** true, at the very least because such entities cannot have an independent existence, being artifacts of an observer's limitations, then **(NN)** must be false. An analogy here might be useful. When dealing with Newtonian gravity, most philosophers would agree that absolute velocities are not part of the fundamental ontology of the theory, but are at most observer-dependent artifacts stemming from a particular choice of reference frame. As a consequence of this fact, any phenomenon/entity that depends for its existence on there being an absolute velocity should be considered an observer-dependent artifact and not be granted the status of an independently existing entity that can enter into

[11] Which entity enters into this condition will depend on one's preferred ontology for spacetime theories. For ease of exposition, I will stick with spacetime points, though analogous arguments can be formulated for any other spacetime ontology one adopts.

[12] Observer-dependent because, as mentioned above, the complexity theory constraints which allow the semiclassical approximation and thus GR to make sense are constraints on an observer's concrete ability to act on and measure a system.

proper explanations and metaphysical constructions.[13] The status of GR in the context of operational recovery is similar, in that GR making sense is an observer-dependent matter. In principle, GR does not make sense, since there are non-zero overlaps between semiclassically distinct states, but since a realistic observer cannot detect such overlaps, they will make observations and have an experience consistent with a world ruled by GR.

Hence, when dealing with operational recovery, a theory of laws involving nomological necessity appears to be too strong a requirement to account for the lawful status of GR as a special science. Not because nomologically necessary connections are in some sense inappropriate in this context, but rather because we lack appropriate entities between which such connections might hold.

6.3 Objections

There are three natural ways to resist the argument presented above. Either retort that the requirement that GR, and more generally special science laws, count as laws is unjustified, or suggest that the kind of observer-dependence of spacetime in operational recovery is benign, or finally, suggest that the specific physical models discussed are too specific and excessively peculiar.

[13] Unless those constructions are allowed to be similarly observer-dependent, which however seems to go against the spirit, if not the letter, of approaches to laws of nature involving nomological necessity.

Let me consider each in turn.

6.3.1 GR Does Not Have Laws

The first objection effectively appeals to the arguments discussed in the literature against the lawful status of special science laws (Woodward, 2000; Cartwright et al., 2005). It uses them to suggest that the requirement for GR to involve laws when seen as special science in QG is excessive, and thus that the discussion in the previous section does not raise a particular problem for theories of laws involving nomological necessity. As I have mentioned before, these kinds of arguments appear to fly in the face of normal scientific practice, which is perfectly happy to countenance special science laws and thus should be treated with great care from the naturalist perspective that should animate a metaphysics of laws.[14] Moreover, as remarked for example by Cohen and Callender (2009), these kinds of arguments are an excessive response to the differences between special and fundamental science laws. While the latter are exceptionless and do not have domain limitations, the former do, and in this sense, one might worry that they are not the same kind of things. However, as Cohen and Callender (2009) argue, the correct response to this state of affairs would be to develop a conception of laws flexible enough to accommodate both types of laws, not to banish special science laws, contra scientific practice.

This observation becomes particularly relevant in the present QG context,

[14] And that certainly animates the discussion on laws of nature in thischapter.

where GR is a low energy effective field theory for QG.[15] Indeed, effective field theories, despite having limited scope because they are valid only up to a certain energy scale, have been long considered, by both physicists and philosophers, perfectly well-defined theories for a scientific realist analysis. Such an analysis would presumably involve a commitment to their lawful content. Indeed, some (Fraser, 2018; Williams, 2020) have even suggested that renormalization group methods[16] should be used to *define* scientific realism. It would be hard to use the renormalization group to define scientific realism if it forbade a commitment to theories having laws of nature. Hence, insofar as GR in the context of QG is an effective field theory, it seems especially natural, even beyond scientific practice considerations, to count it as involving laws.

6.3.2 Approximately Defined Entities Are Enough

The second objection wants to suggest that the kind of ill-defineteness of spacetime highlighted in the previous section is not a threat to theories of laws involving nomological necessity. The objection would run as follows: special science entities are in general defined approximately (think of biological entities, which do not have an exact description in terms of microphysics).

[15] See Wallace (2006) for a philosophical discussion of effective field theories, and Wallace (2021) for this perspective on GR and semiclassical gravity.

[16] The renormalization group is a *semigroup* (group without inverses) acting on the space of possible Lagrangians in terms of which a certain QFT could be defined. Its transformations relate higher energy theories to lower energy theories by a process analogous to coarse-graining, and it constitutes the basic theoretical architecture behind modern talk of effective field theory. See Srednicki (2007) for review.

Since special science entities are approximately defined, we should not place too strict a demand on the conditions for their existence. In particular, the kind of approximations required to recover spacetime in the operational recovery scenario discussed here should be seen as harmless, since it is just another instance of this more general theme.

Let me argue why the scenario under discussion here is different from the kind of situation that merges into something like biology. In doing so, I will also reinforce the claim made in the introduction that operational recovery plays a crucial role in this argument, and hence that there is a concrete reason to focus on this case to explore the full space of possibilities.

In the context of biology, for example, the approximation involved between microphysics and biological entities is in general not supposed to be an observer-dependent matter. Indeed, insofar as one wants to be a realist regarding biological entities and processes, it better not be the case that they are so, for otherwise, our realist commitments would be hard to justify. Rather, what seems to be going on in the case of biology, insofar as there is approximation involved in such cases, is that in general there is no specific collection of particles that at any time defines a certain biological entity. Nonetheless, it is not an observer-dependent fact whether a certain biological entity exists. Indeed, there is no need to invoke the notion of an observer in the explanation I have just given. Biological entities are defined up to a certain level of grain, and not for any finer grain, and this is fine. We do not need to introduce an observer, and we can instead appeal to various objective scales to explain this coarse-graining.

This move is not possible in the cases of operational recovery under discussion here, for in these cases the notion of computational complexity plays a crucial role in defining coarse-graining, and this notion does indeed depend on the observer. It depends on the observer, as I have mentioned before, in the sense that it puts in practice constraints on what an observer can do and measure, not in principle constraints on what is and is not possible. Hence, the two cases are deeply different, and the kind of approximation involved in the context of operational recovery cannot be reduced to the kind of approximation involved in, for example, biology.

6.3.3 Too Much AdS/CFT

As a last possible retort, one might wish to turn the previous observation on the necessity of operational recovery around and argue that this is just a very peculiar example, and so we should not trust it excessively. In particular, while operational recovery is in principle a general thesis, the objection starts from the observation that thus far it has only been applied in the context of AdS/CFT, as discussed in chapter 5, and indeed the structure of AdS/CFT seems to mesh particularly naturally with operational recovery, as explained above. Together, these facts suggest at least a tight relationship between operational recovery and AdS/CFT. However, since AdS/CFT is a controversial physical thesis, this relationship is problematic for operational recovery, or so the objection goes.

To this objection, I wish to give two answers. First, recall that the broad context of this argument is that of naturalized metaphysics (Ladyman et al., 2007) according to which metaphysical theses should be grounded in scientific theories and whose truth should depend, when possible, on the discovery of specific empirical data, as for normal scientific theories. From this point of view, the fact that the argument of this chapter, as it stands, depends on the truth of certain disputed physical theses such as AdS/CFT and holography more generally, is a desirable outcome, not a problem. In this way, insofar as the argument presented here is correct, one can identify an empirical consequence of certain views on laws of nature, rendering them naturalistically acceptable. Hence, the charge that the discussion thus far has depended on controversial physical assumptions should be seen, from a naturalistic perspective, as an advantage. Indeed, in a situation like QG, where there is no common, definitive theoretical framework, it seems that from a naturalist perspective the most one can do is tease out the relationships between different metaphysical theses and various, possibly empirically testable, QG proposals. From this perspective, then, a degree of model dependence is inevitable and should be welcomed. This chapter accomplishes this task for the specific case of operational recovery in AdS/CFT and special science laws.

A second reply to the objection raised above is that, while it is true that the physical theories from which operational recovery was inspired, AdS/CFT and its treatment of black holes, are controversial, the specific results under discussion are more general than the objection seems to recognize. In particular, the basic structure required for operational recovery and the present argument can be derived directly from the gravitational path integral (Penington et al., 2019; Almheiri et al., 2020) without any need for direct holographic input. As such, these features can be seen not as peculiar features of a particular approach to QG, but instead rather general features that emerge directly from a semiclassical analysis of the gravitational path integral. Hence, the charge of excessive reliance on AdS/CFT, even if for the sake of argument taken as legitimate, loses most of its bite. The relevant results for the present chapter are much more general than simply AdS/CFT.

6.4 Humeanism and Operational Recovery

In this section, I will examine whether Humean approaches to laws can deliver the lawful status of GR in QG, contrary to their more modally committed competitors. To illustrate how Humean approaches can account for GR laws when operational recovery holds, I will focus on a particular Humean account of special science laws, the Better Best Systems account of laws (Cohen and Callender, 2009; Callender and Cohen, 2010), and show how it is compatible with operational recovery.

If approaches involving nomological necessity are characterized by their obeying **(NN)**, likewise Humean approaches are characterized by their obeying **(HD)**. Hence, Humean approaches deny the existence of necessary connections between distinct entities, and in particular, deny that there is such a thing as nomological necessity which is encoded by the laws. Nonetheless, Humeans are not eliminativists about laws of nature, i.e., they do not think that they should be removed from our ontologies. Rather, they think that laws of nature should be analyzed in terms of non-modal matters of fact.

The most popular, and in some sense standard, Humean approach to the analysis of laws of nature is the Mills-Ramsey-Lewis (MLR) Best Systems Account (BSA) of laws (Ramsey and Mellor, 1978; Lewis, 1983, 1973; Earman, 1984). According to the BSA, laws of nature are the axioms and theorems of the formal system which encapsulates all facts about the world while maintaining an optimal balance between strength and simplicity. The satisfaction of **(HD)** in this context is immediate. Since according to the BSA laws are, roughly, efficient summaries of matters of fact occurring in the actual world, there is no unanalyzed/primitive modality involved in its account of laws. Hence, there is no necessary connection between distinct entities, as **(HD)** demands. Laws do not give modally charged relations between entities, but rather efficient summaries of what happens in the world.

A non-trivial limitation of the BSA, as emphasized in Callender and Cohen (2010), is that it is unclear how to account for special science laws within it. In particular, special science laws will hardly count as the best summary of the occurrent facts at the actual world, given their limited scope, and so will lose out the title of best system to more fundamental candidates, such as the laws of fundamental physics. Callender and Cohen (2010)'s suggestion to amend this issue is to move to what they call the Better Best System

(BBS) account of laws.[17] According to BBS, laws of nature are the axioms and theorems of the formal system encoding facts about the world that best balances strength and simplicity, *relative to a choice of predicates* in terms of which the systems have to be formulated.

Within BBS, one accounts for special science by choosing an appropriate set of predicates that are relevant to the special science in question and then applying the usual Best Systems algorithm of looking for the best balance of strength and simplicity. The resulting best system is the best system relative to that choice of predicates, and encodes the laws of nature of a world that is partitioned using those predicates. Upon a different choice of predicates, one will find a different best system and hence a different set of laws, so BBS can account for different special and fundamental sciences and count them all as lawful, without rendering the notion of law trivial.

Having said this about Humeanism regarding laws in general, and about BBS in particular, let me now move to consider their relationship with operational recovery. First, let me observe that the problems highlighted in the previous section with the satisfaction of **(NN)** in operational recovery do not apply to the satisfaction of **(HD)**. Indeed, while **(NN)** requires both that there are

[17] Cohen and Callender (2009) originally introduce the BBS to deal with the *problem of immanent comparisons*. Since comparisons of strength and simplicity are always relative to a choice of predicates, it seems that the BSA should provide us with such a recipe (for Lewis, these are the *natural properties*). BBS's solution is instead, as mentioned in the main text, to accept that for each choice of predicates there is a different Best System, and thus that different Best Systems can coexist in the same world. Nonetheless, already in Cohen and Callender (2009), the possibility of using BBS to account for special science laws is noted as an important feature.

distinct entities and that there is a necessary connection between them, for **(HD)** to be satisfied it is sufficient either that the distinct entities in question do not exist, or that there is no necessary connection between them. Hence, the discussion in the previous section not only shows that **(NN)** cannot be satisfied when operational Emergence governs the relation between GR and QG, but it also doubles as an argument that in those same situations **(HD)** is satisfied.[18]

However, this abstract argument has somewhat little bite until we have a concrete Humean proposal for how GR counts as lawful in the context of operational recovery. Here BBS comes to the rescue. To see how GR can count as the best system relative to a specific set of predicates, we, first, need a choice of predicates. Given that we are working in the context of operational recovery, the natural choice is a set of predicates describing the possible observations of a realistic observer, i.e., one capable of performing only operations of subexponential complexity.[19] We know, by the discussion in section §6.1, that an observer limited to operations of subexponential

[18]As one would expect, given that **(NN)** is the negation of **(HD)**.

[19]One might worry that this set of predicates is too ill-defined, along the lines of the arguments made in section §6.2, to appear in BBS. However, as argued in Cohen and Callender (2009), whom I follow here, the most natural way to formulate BBS includes a form of *explosive realism* (Carnap, 1950; Quine, 1969; Kitcher, 2003; Dorr, 2005) about the kinds/predicates that are admissible in BBS. Explosive realism is especially permissive in this context, and in particular, it allows observational predicates. Indeed, the use of an observational language is one of the examples that Cohen and Callender (2009) themselves make to illustrate BBS. Given moreover that any kind of observational language will come with conditions on what is observational, and that these restrictions will often involve practical constraints on what real observers can measure, my choice of predicates is perfectly compatible with BBS.

complexity will make observations and experience a world that is indistinguishable from that of GR (or semiclassical gravity). From this observation, it is natural to derive that insofar as we limit ourselves to predicates that only describe operations of subexponential complexity, then the best system to capture the facts that are relevant to the present situation, relative to this choice of predicates, will be that of GR. Indeed, one can see the claim that a realistic observer's experience would be that of a world governed by GR as a claim that, given a limitation to subexponential operations, they would postulate GR laws as the laws of their world. Given that the limitation to subexponential operations is equivalent to restricting oneself to using predicates describing the possible observations of an observer only capable of subexponentially complex operations, and that the act of postulating a certain set of laws is equivalent to identifying them as the best system[20] (for that set of predicates) then it follows that GR is the best system for the set of predicates describing the observations of an observer limited to subexponentially complex operations. We have thus ensured that GR, despite being a special science in QG, still involves full-blown laws of nature, though of a Humean kind.

[20]This equivalence is the famous epistemological advantage that Humeanism holds over rival theories of laws. See Earman and Roberts (2005) for a sophisticated formulation.

6.5 Conclusions

We have seen how operational recovery is compatible with **(HD)**, the basic condition behind Humeanism about laws. Moreover, we have also seen how BBS, one of the main approaches to Humeanism about laws, can indeed provide an account of GR as a special science in theories of QG interpreted via operational recovery.

Furthermore, we have seen how approaches to laws involving more robust notions of modality such as nomological necessity, characterized via **(NN)** are incompatible with these same scenarios. In this way, I have found a possible, concrete empirical test for theories of laws, in the spirit of naturalized metaphysics. If QG theories, such as AdS/CFT, which naturally involve notions like operational recovery, turn out to be right, then theories of laws involving nomological necessity would be empirically falsified, according to the argument of this chapter. On the other hand, Humean theories of laws such as BBS would receive confirmation.

Chapter 7

Conclusions

Throughout this thesis, I have discussed the foundations and metaphysical implications of recent work on quantum black holes in AdS/CFT (Penington, 2020; Almheiri et al., 2019). Overall, these constructions appear to modify the semiclassical picture of spacetime in subtle and non-trivial ways, to the point of avoiding certain apparently problematic results in GR, and various paradoxes that emerge in the study of quantum black holes, such as the AMPS paradox. Moreover, these constructions appear to lead to fascinating metaphysical consequences: first, they illustrate a formulation of Humeanism compatible with QG, and moreover appear to favor a broadly Humean approach to special science laws, especially if one wants to preserve the lawful status of GR. Moreover, these AdS/CFT models of quantum black holes open the door to strikingly new options for debates around spacetime emergence. In particular, I have argued that spacetime emergence does not make sense for these models, and that spacetime can be recovered from the fundamental degrees of freedom of AdS/CFT only in terms of its operational content; no notion of what spacetime is more ontologically demanding is compatible with these AdS/CFT constructions of black holes.

The extent to which these foundational and metaphysical discussions are tenable is of course open to challenge, and ultimately whether any of this is relevant to the actual world will depend on the success of the attempt at constructing an holographic theory of gravity for cosmological spacetimes like de Sitter. Nonetheless, I take it that the discussion of this thesis has shown that much can learned from a careful study of AdS/CFT, and in particular of models of quantum black holes constructed within this duality. Until we will have a definitive theory of QG, this kind of theory dependent insights will be all that we can have. However, this fact should not be taken as condemning work on the foundations and metaphysics of QG to irrelevance. Rather, in the spirit of naturalized metaphysics and similar lines of inquiry, we should welcome this theory dependence as a way to empirically test our philosophical theories, and thus increase their epistemological standing. My goal throughout this thesis has been to do this, and in this way to illuminate various important conceptual aspects regarding AdS/CFT. Only time will tell how successful my arguments are.

Having said this, many open questions still remain. First, it is far from clear that the non-locality involved in the constructions I have studied is not problematic. A task of extreme importance is understanding whether this non-locality leaks in the low energy regime Marolf and Polchinski (2016). Moreover, it is also essential to understand what it means for locality to be emergent.

Second of all, I have limited myself in this work to the AMPS paradox. However, AMPS is not the only paradox involved in constructing the interior of the black hole. Most notably, in the context of AdS/CFT, I have not discussed the AMPSS paradox Almheiri et al. (2013)[1] and related problems in the field of holographic interior reconstruction Harlow (2018).

Moreover, it would be interesting to know whether there is a relation between the argument in chapter 4 regarding Humeanism and scattering amplitudes in QG and the arguments made in McKenzie (2014) regarding Humeanism in QFT, and whether renormalization group and symmetry arguments might play a role in the present QG discussion.

Finally, it would be interesting to explore the connection of ideas like operational recovery with, for example, the kind of functionalism proposed in Chalmers (2021), to better express the operational characterization of spacetime. Moreover, the idea of operational recovery hints at fascinating consequences for the relevance of global spacetime structure in GR. Lastly, an interpretation of state-dependent operators capable of putting their foundations on the same grounds as the standard linear operators of quantum mechanics seems to be crucial at this point. Investigating all these various aspects is left for future work.

[1] Here AMPSS refers to a different black hole paradox from AMPS.

Bibliography

Akers, C., N. Engelhardt, D. Harlow, G. Penington, and S. Vardhan (2022). The black hole interior from non-isometric codes and complexity. *arXiv preprint arXiv:2207.06536*.

Akers, C. and G. Penington (2021). Leading order corrections to the quantum extremal surface prescription. *Journal of High Energy Physics 2021*(4), 1–73.

Akers, C. and G. Penington (2022). Quantum minimal surfaces from quantum error correction. *SciPost Physics 12*(5), 157.

Almheiri, A., X. Dong, and D. Harlow (2015). Bulk locality and quantum error correction in ads/cft. *Journal of High Energy Physics 2015*(4), 1–34.

Almheiri, A., N. Engelhardt, D. Marolf, and H. Maxfield (2019). The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole. *Journal of High Energy Physics 2019*(12), 1–47.

Almheiri, A., T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini

(2020). Replica wormholes and the entropy of hawking radiation. *Journal of High Energy Physics 2020*(5), 1–42.

Almheiri, A., T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini (2021). The entropy of hawking radiation. *Reviews of Modern Physics 93*(3), 035002.

Almheiri, A., R. Mahajan, J. Maldacena, and Y. Zhao (2020). The Page curve of Hawking radiation from semiclassical geometry. *JHEP 03*, 149.

Almheiri, A., D. Marolf, J. Polchinski, D. Stanford, and J. Sully (2013). An apologia for firewalls. *Journal of High Energy Physics 2013*(9), 1–32.

Almheiri, A., D. Marolf, J. Polchinski, and J. Sully (2013). Black holes: complementarity or firewalls? *Journal of High Energy Physics 2013*(2), 1–20.

Ammon, M. and J. Erdmenger (2015). *Gauge/gravity duality: Foundations and applications*. Cambridge University Press.

Anderson, A. and B. DeWitt (1986). Does the topology of space fluctuate? *Foundations of Physics 16*(2), 91–105.

Armstrong, D. M. (1983). *What is a Law of Nature?* Cambridge, United Kingdom: Cambridge University Press.

Bain, J. (2020). Spacetime as a quantum error-correcting code? *Studies in*

History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 71, 26–36.

Balasubramanian, V., A. Kar, and T. Ugajin (2021). Islands in de sitter space. *Journal of High Energy Physics 2021*(2), 1–32.

Baron, S. (2020). The curious case of spacetime emergence. *Philosophical Studies 177*(8), 2207–2226.

Baron, S. (2021). Eliminating spacetime. *Erkenntnis*, 1–20.

Baron, S. and B. Le Bihan (2022a). Composing spacetime. *The Journal of Philosophy 119*(1), 33–54.

Baron, S. and B. Le Bihan (2022b). Quantum gravity and mereology: Not so simple. *The Philosophical Quarterly 72*(1), 19–40.

Barrett, J. A. (1996). Empirical adequacy and the availability of reliable records in quantum mechanics. *Philosophy of Science 63*(1), 49–64.

Barrett, J. A. (1999). *The quantum mechanics of minds and worlds.* OUP Oxford.

Beem, J. K., P. E. Ehrlich, and K. L. Easley (2017). *Global lorentzian geometry.* Routledge.

Belot, G., J. Earman, and L. Ruetsche (1999). The hawking information loss paradox: The anatomy of controversy. *The British journal for the philosophy of science 50*(2), 189–229.

Berenstain, N. and J. Ladyman (2012). Ontic structural realism and modality. In *Structural Realism*, pp. 149–168. Springer.

Berges, J. (2004). Introduction to nonequilibrium quantum field theory. In *AIP Conference Proceedings*, Volume 739, pp. 3–62. American Institute of Physics.

Bird, A. (2007). *Nature's metaphysics: Laws and properties.* OUP Oxford.

Borde, A., H. Dowker, R. Garcia, R. D. Sorkin, and S. Surya (1999). Causal continuity in degenerate spacetimes. *Classical and Quantum Gravity 16*(11), 3457.

Bousso, R. and G. Penington (2022). Entanglement wedges for gravitating regions. *arXiv preprint arXiv:2208.04993*.

Bousso, R. and E. Wildenhain (2022). Islands in closed and open universes. *Physical Review D 105*(8), 086012.

Brahma, S. (2017). Emergence of time in loop quantum gravity. *Beyond spacetime: The foundations of quantum gravity*, 53–78.

Brennan, T. D., F. Carta, and C. Vafa (2017). The string landscape, the swampland, and the missing corner. *arXiv preprint arXiv:1711.00864*.

Bricker, P. (1996). Isolation and unification: The realist analysis of possible worlds. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 84*(2/3), 225–238.

Brown, A. R., H. Gharibyan, S. Leichenauer, H. W. Lin, S. Nezami, G. Salton, L. Susskind, B. Swingle, and M. Walter (2019). Quantum gravity in the lab: teleportation by size and traversable wormholes. *arXiv preprint arXiv:1911.06314*.

Brown, A. R., H. Gharibyan, G. Penington, and L. Susskind (2020). The python's lunch: geometric obstructions to decoding hawking radiation. *Journal of High Energy Physics 2020*(8), 1–53.

Butterfield, J. (2011a). Emergence, reduction and supervenience: A varied landscape. *Foundations of physics 41*(6), 920–959.

Butterfield, J. (2011b). Less is different: Emergence and reduction reconciled. *Foundations of physics 41*(6), 1065–1135.

Butterfield, J. (2018). On dualities and equivalences between physical theories. *arXiv preprint arXiv:1806.01505*.

Calabrese, P. and J. Cardy (2009). Entanglement entropy and conformal field theory. *Journal of physics a: mathematical and theoretical 42*(50), 504005.

Callender, C. and J. Cohen (2010). Special sciences, conspiracy and the better best system account of lawhood. *Erkenntnis 73*(3), 427–447.

Calzetta, E. A. and B.-L. B. Hu (2008). Nonequilibrium quantum field theory. *Nonequilibrium Quantum Field Theory*.

Carnap, R. (1950). Empiricism, semantics, and ontology. *Revue internationale de philosophie*, 20–40.

Carroll, S. M. (2019). *Spacetime and geometry*. Cambridge University Press.

Cartwright, N. (1989). *Nature's Capacities and their Measurement*. Clarendon Press.

Cartwright, N., A. Alexandrova, S. Efstathiou, A. Hamilton, and I. Muntean (2005). Laws. In M. Smith and F. Jackson (Eds.), *Oxford Handbook of Contemporary Philosophy*. OUP.

Chalmers, D. (2021). Finding space in a nonspatial world. *Philosophy Beyond Spacetime: Implications from Quantum Gravity*, 154–181.

Chen, Y., V. Gorbenko, and J. Maldacena (2021). Bra-ket wormholes in gravitationally prepared states. *Journal of High Energy Physics 2021*(2), 1–61.

Cinti, E., A. Corti, and M. Sanchioni (2022). On entanglement as a relation. *European Journal for Philosophy of Science 12*(1), 1–29.

Cinti, E. and M. Sanchioni (2021a). The devil in the (implicit) details. *International Journal of Theoretical Physics 60*(9), 3234–3253.

Cinti, E. and M. Sanchioni (2021b). Humeanism in light of quantum gravity. *Synthese 199*(3), 10839–10863.

Cohen, J. and C. Callender (2009). A better best system account of lawhood. *Philosophical Studies 145*(1), 1–34.

Colin-Ellerin, S., X. Dong, D. Marolf, M. Rangamani, and Z. Wang (2021a). Real-time gravitational replicas: Formalism and a variational principle. *Journal of High Energy Physics 2021*(5), 1–40.

Colin-Ellerin, S., X. Dong, D. Marolf, M. Rangamani, and Z. Wang (2021b). Real-time gravitational replicas: Low dimensional examples. *Journal of High Energy Physics 2021*(8), 1–54.

Curiel, E. (2019). Singularities and black hole. *Stanford Encyclopedia of Philosophy*.

Cushing, J. T. (1990). *Theory construction and selection in modern physics: The S matrix*. Cambridge University Press.

Darby, G. (2009). Lewis's worldmate relation and the apparent failure of humean supervenience. *dialectica 63*(2), 195–204.

Darby, G. (2020). Relational holism and humean supervenience. *The British Journal for the Philosophy of Science*.

De Haro, S., D. R. Mayerson, and J. N. Butterfield (2016). Conceptual aspects of gauge/gravity duality. *Foundations of Physics 46*(11), 1381–1425.

DeWitt, C. and B. DeWitt (1964). Relativity, groups and topology. *Relativité*.

Dijkgraaf, R., B. Heidenreich, P. Jefferson, and C. Vafa (2018). Negative branes, supergroups and the signature of spacetime. *Journal of High Energy Physics 2018*(2), 1–63.

Dorr, C. (2005). What we disagree about when we disagree about ontology. *Fictionalism in metaphysics 234*, 86.

Dorr, C. and J. Hawthorne (2013). Naturalness. *Oxford Studies in Metaphysics, Volume 8*, 1.

Dowker, F. and S. Surya (1998). Topology change and causal continuity. *Physical Review D 58*(12), 124019.

Earman, J. (1984). Laws of nature: The empiricist challenge. In *DM Armstrong*, pp. 191–223. Springer.

Earman, J. (2008). Pruning some branches from "branching spacetimes". *Philosophy and Foundations of Physics 4*, 187–205.

Earman, J. (2015). Some puzzles and unresolved issues about quantum entanglement. *Erkenntnis 80*(2), 303–337.

Earman, J. (2021). Implementing david lewis' principal principle: A program for investigating the relation between credence and chance. *http://philsci-archive.pitt.edu/19922/*.

Earman, J. and J. T. Roberts (2005). Contact with the nomic: A challenge for deniers of humean supervenience about laws of nature part ii: The epistemological argument for humean supervenience. *Philosophy and Phenomenological Research 71*(2), 253–286.

Earman, J. and G. Valente (2014). Relativistic causality in algebraic quantum field theory. *International Studies in the Philosophy of Science 28*(1), 1–48.

Einstein, A., B. Podolsky, and N. Rosen (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev. 47*, 777–780.

Einstein, A. and N. Rosen (1935). The particle problem in the general theory of relativity. *Phys. Rev. 48*, 73–77.

Engelhardt, N., G. Penington, and A. Shahbazi-Moghaddam (2021). A world without pythons would be so simple. *Classical and Quantum Gravity 38*(23), 234001.

Engelhardt, N., G. Penington, and A. Shahbazi-Moghaddam (2022). Finding pythons in unexpected places. *Classical and Quantum Gravity 39*(9), 094002.

Fraser, J. D. (2018). Renormalization and the formulation of scientific realism. *Philosophy of Science 85*(5), 1164–1175.

French, S. (2014). *The structure of the world: Metaphysics and representation.* Oxford University Press.

Frigg, R. (2007). A field guide to recent work on the foundations of thermodynamics and statistical mechanics. *The Ashgate companion to the new philosophy of physics*, 99–196.

Geroch, R. (1970). Singularities. In *Relativity*, pp. 259–291. Springer.

Gibbons, G. W. and S. W. Hawking (1993). Action integrals and partition functions in quantum gravity. In *Euclidean Quantum Gravity*, pp. 233–237. World Scientific.

Gibbons, G. W., S. W. Hawking, and M. J. Perry (1978). Path integrals and the indefiniteness of the gravitational action. In *Euclidean quantum gravity*, pp. 102–111. World Scientific.

Giddings, S. B. (2015). Hilbert space structure in quantum gravity: an algebraic perspective. *Journal of High Energy Physics 2015*(12), 1–21.

Glick, D. and G. Darby (2018). In defense of the metaphysics of entanglement. *The Foundation of Reality: Fundamentality, Space, and Time*, 169–180.

Goto, K., T. Hartman, and A. Tajdini (2021). Replica wormholes for an evaporating 2d black hole. *Journal of High Energy Physics 2021*(4), 1–57.

Green, M. B., J. H. Schwarz, and E. Witten (2012). *Superstring theory.* Cambridge university press.

Haag, R. (1992). *Local quantum physics: Fields, particles, algebras*. Springer Science & Business Media.

Halliwell, J. J. and J. B. Hartle (1990). Integration contours for the no-boundary wave function of the universe. *Physical Review D 41*(6), 1815.

Halvorson, H. (2007). Algebraic quantum field theory. In J. Earman and J. Butterfield (Eds.), *Handbook of Philosophy of Physics*, pp. 731–922. Elsevier.

Harlow, D. (2014). Aspects of the papadodimas-raju proposal for the black hole interior. *Journal of High Energy Physics 2014*(11), 1–39.

Harlow, D. (2016). Jerusalem lectures on black holes and quantum information. *Reviews of Modern Physics 88*(1), 015002.

Harlow, D. (2017). The ryu–takayanagi formula from quantum error correction. *Communications in Mathematical Physics 354*(3), 865–912.

Harlow, D. (2018). TASI Lectures on the Emergence of Bulk Physics in AdS/CFT. *PoS TASI2017*, 002.

Hartle, J. B. and S. W. Hawking (1983). Wave function of the universe. In *Euclidean Quantum Gravity*, pp. 310–325. World Scientific.

Hartman, T., Y. Jiang, and E. Shaghoulian (2020). Islands in cosmology. *Journal of High Energy Physics 2020*(11), 1–56.

Hau, L. A., J. L. F. Dorado, and M. S. Caja (2020). Structure of globally hyperbolic spacetimes-with-timelike-boundary. *Revista matemática iberoamericana 37*(1), 45–94.

Hawking, S. W. (1976). Breakdown of predictability in gravitational collapse. *Physical Review D 14*(10), 2460.

Hawking, S. W. (1978). Spacetime foam. *Nuclear Physics B 144*(2-3), 349–362.

Hayden, P. and G. Penington (2019). Learning the alpha-bits of black holes. *Journal of High Energy Physics 2019*(12), 1–55.

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

Huggett, N. and C. Wüthrich (2013). 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.

Iqbal, A., C. Vafa, N. Nekrasov, and A. Okounkov (2008). Quantum foam and topological strings. *Journal of High Energy Physics 2008*(04), 011.

Jackiw, R. (1985). Lower dimensional gravity. *Nuclear Physics B 252*, 343–356.

Jafferis, D., A. Zlokapa, J. D. Lykken, D. K. Kolchmeyer, S. I. Davis, N. Lauk,

H. Neven, and M. Spiropulu (2022). Traversable wormhole dynamics on a quantum processor. *Nature 612*(7938), 51–55.

Jaksland, R. (2021). Entanglement as the world-making relation: Distance from entanglement. *Synthese 198*(10), 9661–9693.

Kitcher, P. (2003). *Science, truth, and democracy*. Oxford University Press.

Kodama, H. (1979). Inevitability of a naked singularity associated with the black hole evaporation. *Progress of Theoretical Physics 62*(5), 1434–1435.

Kontsevich, M. and G. Segal (2021). Wick rotation and the positivity of energy in quantum field theory. *The Quarterly Journal of Mathematics 72*(1-2), 673–699.

Krasnikov, S. (2014). Corrigendum: No time machines in classical general relativity (2002 class. quantum grav. 19 4109). *Classical and Quantum Gravity 31*(7), 079503.

Ladyman, J., D. Ross, J. Collier, and D. Spurrett (2007). *Every thing must go: Metaphysics naturalized*. Oxford University Press on Demand.

Lam, V. and C. Wüthrich (2018). Spacetime is as spacetime does. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 64*, 39–51.

Lam, V. and C. Wüthrich (2021a). Laws beyond spacetime. *arXiv preprint arXiv:2111.04500*.

Lam, V. and C. Wüthrich (2021b). Spacetime functionalism from a realist perspective. *Synthese 199*(2), 335–353.

Le Bihan, B. (2020). String theory, loop quantum gravity and eternalism. *European Journal for Philosophy of Science 10*(2), 1–22.

Le Bihan, B. (2021). Spacetime emergence in quantum gravity: Functionalism and the hard problem. *Synthese 199*(2), 371–393.

Lesourd, M. (2018). Causal structure of evaporating black holes. *Classical and Quantum Gravity 36*(2), 025007.

Lewis, D. (1973). *Counterfactuals*. John Wiley & Sons.

Lewis, D. (1983). New work for a theory of universals. *Australasian journal of philosophy 61*(4), 343–377.

Lewis, D. (1986). *On the plurality of worlds*, Volume 322. Blackwell Oxford.

Lewis, D. (1994). Humean supervenience debugged. *Mind 103*(412), 473–490.

Linnemann, N. (2021). On the empirical coherence and the spatiotemporal gap problem in quantum gravity: and why functionalism does not (have to) help. *Synthese 199*(2), 395–412.

Loewer, B. (1996). Humean supervenience. *Philosophical Topics 24*(1), 101–127.

Louko, J. and R. D. Sorkin (1997). Complex actions in two-dimensional topology change. *Classical and Quantum Gravity 14*(1), 179.

Malament, D. B. (1977). The class of continuous timelike curves determines the topology of spacetime. *Journal of mathematical physics 18*(7), 1399–1404.

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

Maldacena, J. (1999). The large-n limit of superconformal field theories and supergravity. *International journal of theoretical physics 38*(4), 1113–1133.

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

Maldacena, J., G. J. Turiaci, and Z. Yang (2021). Two dimensional nearly de sitter gravity. *Journal of High Energy Physics 2021*(1), 1–60.

Manchak, J. and J. O. Weatherall (2018). (information) paradox regained? a brief comment on maudlin on black hole information loss. *Foundations of Physics 48*(6), 611–627.

Manchak, J. B. (2009). Can we know the global structure of spacetime? *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 40*(1), 53–56.

Manchak, J. B. (2013). Global space time structure. In R. Batterman (Ed.), *The Oxford Handbook of Philosophy of Physics.*

Marolf, D. and H. Maxfield (2020). Transcending the ensemble: baby universes, spacetime wormholes, and the order and disorder of black hole information. *Journal of High Energy Physics 2020*(8), 1–72.

Marolf, D. and H. Maxfield (2021). Observations of hawking radiation: the page curve and baby universes. *Journal of High Energy Physics 2021*(4), 1–79.

Marolf, D. and J. Polchinski (2016). Violations of the Born rule in cool state-dependent horizons. *JHEP 01*, 008.

Matarese, V. (2019). Loop quantum gravity: A new threat to humeanism? part i: The problem of spacetime. *Foundations of Physics 49*(3), 232–259.

Maudlin, T. (1994). *Quantum non-locality and relativity: Metaphysical intimations of modern physics.* John Wiley & Sons.

Maudlin, T. (2007). *The metaphysics within physics.* Oxford University press on demand.

Maudlin, T. (2017). (Information) Paradox Lost. *arXiv preprint arXiv:1705.03541*.

McKenzie, K. (2014). In no categorical terms: A sketch for an alternative route to a humean interpretation of laws. In *New Directions in the Philosophy of Science*, pp. 45–61. Springer.

Milton, J. (1674). *Paradise Lost.* London ; New York : Penguin Books, 2000.

Mussardo, G. (2010). *Statistical field theory: an introduction to exactly solved models in statistical physics.* Oxford University Press.

Ney, A. (2021). *The world in the wave function: a metaphysics for quantum physics.* Oxford University Press.

Nezami, S., H. W. Lin, A. R. Brown, H. Gharibyan, S. Leichenauer, G. Salton, L. Susskind, B. Swingle, and M. Walter (2021). Quantum gravity in the lab: teleportation by size and traversable wormholes, part ii. *arXiv preprint arXiv:2102.01064*.

Nielsen, M. A. and I. L. Chuang (2010). *Quantum Computation and Quantum Information.* Cambridge University Press.

Norton, J. D. (2022). How to make possibility safe for empiricists. In *Rethinking the Concept of Law of Nature: Natural Order in the Light of Contemporary Science*, pp. 129–159. Springer.

O'Connor, T. (2021). Emergent Properties. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021 ed.). Metaphysics Research Lab, Stanford University.

Page, D. (1993a). Average entropy of a subsystem. *Physical review letters 71*(9), 1291.

Page, D. (1993b). Information in black hole radiation. *Physical review letters 71*(23), 3743.

Papadodimas, K. and S. Raju (2013). An infalling observer in ads/cft. *Journal of High Energy Physics 2013*(10), 1–73.

Papadodimas, K. and S. Raju (2016). Remarks on the necessity and implications of state-dependence in the black hole interior. *Physical Review D 93*(8), 084049.

Penington, G. (2020). Entanglement wedge reconstruction and the information paradox. *Journal of High Energy Physics 2020*(9), 1–84.

Penington, G., S. H. Shenker, D. Stanford, and Z. Yang (2019). Replica wormholes and the black hole interior. *arXiv preprint arXiv:1911.11977*.

Penington, G. and E. Witten (2023). Algebras and states in jt gravity. *arXiv preprint arXiv:2301.07257*.

Psillos, S. (2014). *Causation and explanation.* Routledge.

Quine, W. V. (1969). Natural kinds. In *Ontological relativity and other essays*, pp. 114–138. Columbia University Press.

Ramsey, F. P. and D. H. Mellor (1978). *Foundations: Essays in philosophy, logic, mathematics and economics.* Routledge & Kegan Paul London.

Ramsey, W. (2022). Eliminative Materialism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022 ed.). Metaphysics Research Lab, Stanford University.

Rovelli, C. and F. Vidotto (2015). *Covariant loop quantum gravity: an elementary introduction to quantum gravity and spinfoam theory.* Cambridge University Press.

Saad, P., S. H. Shenker, and D. Stanford (2018). A semiclassical ramp in syk and in gravity. *arXiv preprint arXiv:1806.06840*.

Schaffer, J. (2005). Quiddistic knowledge. *Philosophical Studies 123*(1), 1–32.

Shoemaker, S. (1980). Causality and properties. In P. van Inwagen (Ed.), *Time and cause*, pp. 109–135. Springer.

Srednicki, M. (2007). *Quantum field theory.* Cambridge University Press.

Susskind, L. (2012). Singularities, firewalls, and complementarity.

Susskind, L. (2016a). Copenhagen vs Everett, Teleportation, and ER=EPR. *Fortsch. Phys. 64*(6-7), 551–564.

Susskind, L. (2016b). Entanglement is not enough. *Fortschritte der Physik 64*(1), 49–71.

Susskind, L., L. Thorlacius, and J. Uglum (1993, Oct). The stretched horizon and black hole complementarity. *Physical Review D 48*(8), 3743–3761.

Teitelboim, C. (1983). Gravitation and hamiltonian structure in two space-time dimensions. *Physics Letters B 126*(1-2), 41–45.

Tong, D. (2005). Tasi lectures on solitons. *arXiv preprint hep-th/0509216*.

Tong, D. (2009). Lectures on string theory. *arXiv preprint arXiv:0908.0333*.

Unruh, W. G. and R. M. Wald (1995). On evolution laws taking pure states to mixed states in quantum field theory. *Phys. Rev. D 52*, 2176–2182.

Unruh, W. G. and R. M. Wald (2017). Information Loss. *Rept. Prog. Phys. 80*(9), 092002.

Visser, M. (2022). Feynman's $i\epsilon$ prescription, almost real spacetimes, and acceptable complex spacetimes. *Journal of High Energy Physics 2022*(8), 1–28.

Wald, R. M. (1984). Black holes, singularities and predictability. In *Quantum theory of gravity. Essays in honor of the 60th birthday of Bryce S. Dewitt*.

Wallace, D. (2006). In defence of naiveté: The conceptual status of lagrangian quantum field theory. *Synthese 151*(1), 33–80.

Wallace, D. (2020, 4). Why Black Hole Information Loss is Paradoxical. In N. Huggett, K. Matsubara, and C. Wüthrich (Eds.), *Beyond Spacetime*, pp. 209–236. Cambridge University Press.

Wallace, D. (2021). Quantum gravity at low energies. *arXiv preprint arXiv:2112.12235*.

Weatherall, J. O. (2022). Where does general relativity break down? *arXiv preprint arXiv:2204.03869*.

Williams, P. (2020). Scientific realism made effective. *The British Journal for the Philosophy of Science*.

Wilson, A. (2021). Explanations of and in time. In *Philosophy Beyond Spacetime: Implications from Quantum Gravity*, pp. 182–198. Oxford University Press.

Wilson, J. (2010). What is hume's dictum, and why believe it? *Philosophy and Phenomenological Research 80*(3), 595–637.

Wilson, J. (2015). Metaphysical emergence: Weak and strong. *Metaphysics in contemporary physics 251*, 306.

Witten, E. (1998). Anti de sitter space and holography. *Advances in Theoretical and Mathematical Physics 2*(2), 253–291.

Witten, E. (2001). Quantum gravity in de sitter space. Technical report.

Witten, E. (2018). What every physicist should know about string theory. In *Foundations of mathematics and physics one century after Hilbert*, pp. 197–210. Springer.

Witten, E. (2021). A note on complex spacetime metrics. *arXiv preprint arXiv:2111.06514*.

Woodward, J. (2000). Explanation and invariance in the special sciences. *The British journal for the philosophy of science 51*(2), 197–254.

Wüthrich, C. (2017). Raiders of the lost spacetime. In *Towards a theory of spacetime theories*, pp. 297–335. Springer.

Wüthrich, C. (2020). When the actual world is not even possible. In *The Foundation of Reality*, pp. 233–254. Oxford University Press.