Black Hole Paradoxes

A Unified Framework for Information Loss

by

Saakshi Dulani

M.A., Columbia University, 2019
B.S.F.S., Georgetown University, 2013

Submitted to the Graduate Faculty of Humanities
in Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

University of Geneva

December 2023

Copyright © 2024 by Saakshi Dulani

All Rights Reserved

Dedication

To all the giants on whose shoulders I stand: my beloved family, dear friends, cherished mentors, and steadfast partner.

Black Hole Paradoxes: A Unified Framework for Information Loss

Abstract

The black hole information loss paradox is a catch-all term for a family of puzzles related to black hole evaporation. For almost 50 years, the quest to elucidate the implications of black hole evaporation has not only sustained momentum, but has also become increasingly populated with proposals that seem to generate more questions than they purport to answer. Scholars often neglect to acknowledge ongoing discussions within black hole thermodynamics and statistical mechanics when analyzing the paradox, including the interpretation of Bekenstein-Hawking entropy, which is far from settled. To remedy the dialectical gridlock, I have formulated an overarching, unified framework, which I call "Black Hole Paradoxes", that integrates the debates and taxonomizes the relevant 'camps' or philosophical positions.

I demonstrate that black hole evaporation within Hawking's semi-classical framework insinuates how late-time Hawking radiation is an entangled global system, a contradiction in terms. The relevant forms of information loss are associated with a decrease in maximal Boltzmann entropy and an increase in global von Neumann entropy respectively, which engender what I've branded the "paradox of phantom entanglement". Prospective solutions are then tasked with demonstrating how late-time Hawking radiation is either exclusively an entangled subsystem, in which a black hole remnant lingers as an information safehouse, or exclusively an unentangled global system, in which information is evacuated to the exterior.

The disagreement between safehouse and evacuation solutions boils down to the statistical interpretation of thermodynamic black hole entropy, i.e., Bekenstein-Hawking entropy. Safehouse solutions attribute Bekenstein-Hawking entropy to a minority of black hole degrees of freedom, those that are associated with the horizon. Evacuation solutions, in contrast, attribute Bekenstein-Hawking entropy to all black hole degrees of freedom. I argue that the interpretation of Bekenstein-Hawking entropy is the litmus test to vet the overpopulated proposal space. So long as any proposal rejecting Hawking's original calculation independently derives black hole evaporation, globally conserves degrees of freedom and entanglement, preserves a version of semi-classical gravity at sub-Planckian scales, and describes black hole thermodynamics in statistical terms, then it counts as a genuine solution to the paradox.

Keywords: black hole information loss paradox, phantom entanglement, black hole thermodynamics, black hole statistical mechanics, Bekenstein-Hawking entropy, Page-time paradox, guiding principles, semi-classical gravity, quantum gravity

Contents

List of Figures

List of Equations

Preface

Who knew that voraciously reading Susskind's popular book, *The Black Hole War: My Battle with Stephen Hawking to Make the World Safe for Quantum Mechanics*, to cope with the roller-coaster ride of living and working in the magnificent yet intricate country of Türkiye, would change my life's trajectory, catapulting me on a collision course with philosophy of physics. Yet here I am seven years later, looking back on the promise to myself to get to the bottom of the black hole information loss paradox.

First and foremost, I'm deeply indebted to Christian Wüthrich, without whom I would never have arrived at the equally magnificent yet intricate country of Switzerland for my doctoral studies. After watching his clear, concise, and enthralling lecture about black hole entropy at the 2018 "Chimera of Entropy" conference in Croatia, I definitively identified who I wanted as my PhD advisor. He has since been a guiding and stabilizing force in my academic journey, from helping me sharpen my philosophical reasoning skills with plenty of autonomy to figure out my own positions, to helping me navigate my experience abroad. I wish to express my heartfelt gratitude to Chris for his invaluable role as a mentor, and I'm honored to have been his student.

Indeed, my time at the University of Geneva would not have been nearly so enriching if it weren't for the exceptionally talented and kind colleagues that brought to life the department's philosophy of physics groups. I'd like to thank Baptiste le Bihan, a close collaborator and friend, for his sincere engagement with my unconventional ideas. He used his gift for literary metaphors to transform my epithets, such as "phantom entanglement", into careful metaphysical concepts. I'm grateful for the afternoons we spent debating the metaphysics of black holes and drawing Penrose diagrams on the blackboard. I'd also like to thank Claudio Calosi, Lorenzo Cocco, Lucy James, Niels Linneman, Emilia Margoni, Tannaz Najafi, Maria Nørgaard, Marta Pedroni, Tim Riedel, David Schroeren, Albert Solé, and Charlotte Zito for their input and unconditional support.

Outside my circle at the University of Geneva, I wish to thank many others who have extended their generosity. During the COVID-19 pandemic, when the whole world was under lockdown, I was waiting it out in the U.S. I'm extremely grateful to everyone around me who accommodated my needs during this difficult period with understanding and grace. Barry Loewer and Jill North took me under their wing as a remote visiting student at Rutgers University, as did David Wallace at the University of Pittsburgh. In fact, I owe much of my inspiration to David, who patiently sat through numerous calls and email messages to discuss black hole thermodynamics

and statistical mechanics. I ended up writing sizable chunks of my dissertation in an imagined dialogue with him, aspiring to match his finesse and counter that the Page-time paradox isn't the only version of black hole information loss worth pursuing. I only lament that more nuggets of inspiration from those conversations didn't make it into the dissertation, since he has kindly agreed to evaluate it as a committee member, but I will definitely build on those ideas in future projects.

Speaking of future projects, I'm incredibly excited to collaborate with Sean Carroll at Johns Hopkins University. His insights helped me bridge an all-too common divide between physicists and philosophers, informing how I crafted my approach to the black hole information loss paradox. I'm beyond appreciative for his input and interest in my work as a PhD student. I'm also deeply indebted to my academic references, David Albert and Brian Greene (in addition to my advisor, Chris), for their instrumental support in securing me doctoral funding as well as a job after graduation.

Moreover, I'm immensely grateful to Nick Huggett for seeing me to the finish line as his predoctoral fellow and having me polish/present my ideas at his dynamic philosophy of physics group at the University of Illinois-Chicago. The same goes for Karen Crowther, who graciously allowed me to crash her conference at the University of Oslo on theoretical principles. She's another role model for the direction in which I've taken my project, and I'm thankful that she's on my committee. My gratitude also extends to Patricia Palacios, the last but not least of the external expert committee members. During my first week in Geneva, I attended her EPSA talk on the universality of Hawking radiation, and as they say, the rest is history.

This list would be incomplete without acknowledging Tim Maudlin, to whom I'm grateful for engaging me on alternative perspectives, alongside Daniel Sudarsky, for offering me a seat at the table at an invitation-only black hole information loss conference at the John Bell Institute (a perfect occasion to visit Croatia again). I confess that I've greatly enjoyed quoting Tim and can only hope to learn penning my thoughts to paper with as much bold intrigue. I must say the same for Erik Curiel, who has also indulged my questions and offered creative, astute analyses of the discourse. Both Tim and Erik brought unspoken nuances of the debate to my attention, which I aim to have channeled for the greater good in my dissertation.

Finally, I wish to thank everyone with whom I've had constructive conversations on topics related to this project: Craig Callender, Sam Fletcher, Nicolas Gisin, Stephen Hsu, Bernard Kay, Maulik Parikh, Carina Prunkl, Carlo Rovelli, Laura Ruetsche, Julian Sonner, Chris Timpson, Bob Wald, and countless others too numerous to name.

This work has been funded in part by the U.S. Fulbright Program, the Swiss Federal Commission for Scholarships for Foreign Students, the Cogito Foundation, and the John Templeton Foundation for the Beyond Spacetime Predoctoral Fellowship. I have also received travel support from the Ludwig Maximilian University of Munich and the University of Oslo. The financial and administrative assistance of these funding bodies is gratefully acknowledged, though the views expressed in this dissertation are solely those of the author.

Introduction

John Archibald Wheeler, one of the great American physicists of the 20th century, was notorious for coining catchy phrases. It's reverent folklore that he invented the locution 'black hole', which every layperson knows to be an astrophysical object with such immense gravity at its core that not even light cannot escape. Additionally, he came up with the expression "it from bit"; he explains that "every it — every particle, every field of force, even the spacetime continuum itself — derives its function, its meaning, its very existence entirely — even if in some contexts indirectly — from the apparatus-elicited answers to yes or no questions, binary choices, bits" (Wheeler, 2002, p. 309). A prominent challenge in theoretical physics today is to ascertain whether these bits get lost in black holes.

Lured by Wheeler's adage "it from bit," theoretical physicists are increasingly tempted to interpret the foundations of physics as consisting in information ('bit'), rather than substances such as particles or fields ('it'). The consequences of this 'informational turn' in physics are many and profound, including at the frontier of contemporary physics where the question arises whether black holes gobble up information or are the universe's cloud storage.

0.1 Information in Physics: Entropy

Interest in applying information-theoretic infrastructure to black hole physics has exploded over the last couple of decades (see e.g., Bekenstein 2001; Hayden and Preskill 2007; Harlow and Hayden 2013), much to the chagrin of many philosophers of physics, who regard it as a category mistake. Wüthrich (2019), for one, proclaims,

> [F]undamental physics is about the objective structure of our world, not about our beliefs or our information (p. 216).

On the other side of the fence, Bekenstein and Schiffer (1990) defend the physicality of information.

> We take it as axiomatic here that there is no such thing as disembodied information, information in the abstract. Information, of whatever kind, must be associated with matter, radiation, or fields of some sort (p. 355).

Landauer (1996) goes on to emphasize the representational dependence of information on material substrates, and therefore, its submission to fundamental laws of nature.

However, as Floridi (2004) points out, we barely have a basic grasp of what information is, despite knowing how to organize, quantify, compress, transmit, and store it. So, the accusation of a category mistake may be premature.

Another concept in physics has raised similar questions: that of entropy. Is entropy an objective feature of the world (see Albert 2000; Shenker 2019) or a subjective representation of epistemic agents (see Jaynes 1957a; Jaynes 1957b)? Like information, we know how to measure it. In thermodynamics, the study of heat flow, changes in entropy S are driven by changes in heat Q at a characteristic temperature, T:

$$dS = \frac{dQ}{T}. \tag{1}$$

Heat is energy, a physical entity, and temperature is a property of heat, so it follows that entropy is a physical property. Given that the entropy of an isolated system never goes down (barring highly improbable fluctuations accounted for in statistical mechanics), it is commonly thought of as a measure of degraded energy, which overwhelmingly increases over time.

In statistical mechanics, which deals with large systems, entropy appears as various statistical measures over degrees of freedom, which represent the fundamental constituents and their accessible states within the chosen theory. For example, Boltzmann entropy is a counting measure. It enumerates how many microstates are consistent with a system's macroscopic parameters, such as its temperature, volume, and pressure. At first glance, it seems here that entropy is an objective, physical property. For an appropriate choice of coarse-graining, Boltzmann entropy agrees with thermodynamic entropy: macrostates in global thermodynamic equilibrium have many more corresponding microstates than macrostates out of global thermodynamic equilibrium. Following Hawking's approach, non-equilibrium macrostates can be more accurately described as temporarily equilibrated macrostates within a quasi-static equilibration process.

However, to get a statistical account of thermodynamic evolution off the ground and link macrostate degeneracy to a system's likelihood of occupying it, the relationship between entropy and probability becomes indispensable. By imposing a probability distribution over the state space, Gibbs entropy explicates how a system is most likely to occupy a microstate that is a member of the largest macrostate, or if it's not already in that macrostate, how it will almost certainly evolve towards it. Such an interpretation naturally leads to the conclusion that Gibbs entropy is not an objective feature of physical systems at all, but rather, our limited knowledge and degree of control over them, earning it a subjective reputation. In a rather Jaynesian fashion, the macroscopic irreversibility of the Second Law of Thermodynamics can be viewed as an epistemic phenomenon. Jaynes (1957b) states,

> [I]t is not the physical process that is irreversible, but rather our ability to follow it (p. 171).

Coincidentally, Gibbs entropy shares the same mathematical structure as Shannon entropy, an information measure born in the field of communications science (Shan-

non, 1948). Philosophers have thus relegated the concept of information to the realm of epistemic agents (see Maroney and Timpson 2017). Many physicists, on the other hand, view the isomorphism between statistical mechanical entropy and Shannon entropy as highly suggestive of a deep connection between physics and information, including Bekenstein (1973), the founder of black hole entropy (see also Bekenstein 1981; Bekenstein 2007).

I agree with the physicists, and for the purpose of this project, I'm presupposing from the outset a similar thesis to that of Myrvold (2021), one of the few philosophers in the same boat, where statistical mechanical entropy embodies a specific definition of information with both epistemic and objective features. In particular, when Shannon entropy takes on the identity of Gibbs/Boltzmann entropy, the epistemic interpretation of the Second Law as increasing ignorance is connected to objective and physically salient patterns in the ontology characterizing thermodynamic evolution. This position about the twofold nature of entropy holds more water with the advent of quantum entanglement and the dual function of von Neumann entropy as Gibbs and entanglement entropy, the latter of which has pronounced ontological origins (see e.g., Mermin 1998).

0.2 Information in Black Hole Physics: Bekenstein-Hawking Entropy

Insofar as entropy is the bridge between physics and information, the instinctive inference should be that entropy – in some way, shape, or form – is the bedrock of the black hole information loss paradox. But apparently, this connection hasn't been obvious, so the primary aim of my project is to push our understanding of the relationship between information and entropy in the context of black hole physics and illuminate how all roads lead to black hole entropy.

Bekenstein (1973) postulated that black holes have entropy proportional to horizon area to save a Generalized Second Law of Thermodynamics (GSL). Given their relativistic nature as perfect absorbers, black holes seem to be prime candidates for violating the Second Law. They can systematically reduce the universe's entropy simply by consuming matter. However, when black holes consume matter, their mass and surface area expand. He hypothesized that the strict non-decrease in horizon area as mathematically proven by Hawking's area theorem emulates another law in physics: the non-decrease of an isolated system's thermodynamic entropy.

By making use of this parallel, Bekenstein predicted that a black hole's entropy is proportional to its surface gravity, the force exerted from infinitely far away to keep an object stationary at the event horizon, which is also a function of mass/area. Therefore, he concluded that when black holes consume matter, they do not actually rid the universe of its entropic waste. Their increased surface area, and hence entropy, more than compensates for the initial entropic disposal.

Squarely within general relativity though, there are challenges to the conception

of black hole entropy. The Second Law is typically a statement about energy degradation involving heat dissipation, but black holes don't radiate classically. And if thermodynamic entropy is to have statistical mechanical underpinnings, black holes of fixed mass, charge and spin would have numerous possible microscopic configurations, thereby violating the no-hair theorem that limits black hole information to these three macroscopic parameters. Instead, Bekenstein (1973) proposed a non-thermodynamic interpretation of black hole entropy as an information measure over possible histories of formation. But only after Hawking (1975) added quantum field theory to the mix in semi-classical gravity, demonstrating that black holes evaporate at temperatures inversely proportional to their mass, was Bekenstein entropy conceived of as equivalent to thermodynamic entropy and renamed Bekenstein-Hawking entropy (Bekenstein, 1994).[1]

Despite lending credence to the idea of thermodynamic black hole entropy, Hawking radiation has introduced a deep puzzle. As an evaporating black hole shrinks and eventually dissipates, its radiation remains thermal up to the last moment of a black hole's life. In other words, the distribution of radiation energy modes is always independent of the details of the gravitational source, such as the matter that formed the black hole. Hawking predicted that after a black hole evaporates, leaving only radiation in its wake, all details about the original collapsing matter disappear from the universe. If we input data about the final radiation state into the dynamical laws, it would be impossible to recover, even in principle, complete information about prior states. This failure of retrodictability violates global unitary evolution (deterministic quantum dynamics) and has been dubbed information loss (see Hawking 1976).

There's a harrowing question about what this failure of retrodictability entails, one of the startling consequences being that the extent of entanglement, i.e., nonlocal correlations with the environment, has grown for what appears to be an isolated system (see Page 1995). The specious explanation behind this strange evolution is that late-time Hawking radiation is still entangled with matter trapped inside the black hole; however, since that matter could not have crossed the event horizon and radiated away without bypassing the speed-of-light barrier, it must have disappeared at the central singularity.

However, forcibly reinstating retrodictability (even nonlocally) risks undermining the event horizon as a smooth and "drama-free" boundary, since snapping late-time entanglement releases an enormous amount of energy, thus vaporizing any infalling matter at the point of no-return (see Almheiri et al. 2013). Within an influential intellectual current permeating the physics discourse, information loss is also cast as a failure of the equivalence principle in the face of global unitary evolution, and therefore, as a paradigmatic clash between general relativity and quantum theory (see Susskind 2008). With or without global unitary evolution, we seem to face a paradigm-shifting

[1] One might be worried that the Second Law is threatened once more if black hole evaporation reduces the size and mass of the black hole, and therefore, its Bekenstein-Hawking entropy. The increase in entropy of the Hawking radiation, however, outpaces the decrease in the black hole's Bekenstein-Hawking entropy (see Zurek 1982).

paradox.

So, the black hole information loss paradox is a catch-all term for a family of puzzles related to black hole evaporation. For almost 50 years, the quest to elucidate the implications of black hole evaporation has not only sustained momentum, but has also become increasingly populated with proposals that seem to generate more questions than they purport to answer. What I aim to contest in this dissertation is the mainstream narrative surrounding the conceptual breakdown that has unfortunately popularized a red-herring – that indeterminism simpliciter is paradoxical. It has thus far not sufficiently been appreciated in the literature that the controversy surrounding the status and legitimacy of the black hole information loss paradox is infused with an underlying debate over the status and legitimacy of black hole statistical mechanics. To make headway on the paradox, we need to get to the bottom of Bekenstein-Hawking entropy.

Bekenstein-Hawking entropy is perhaps the most mysterious type of entropy due to its revisionary connection between entropy and geometry. In terrestrial physics, we expect a system's entropy to scale with its volume – more space means more available states. However, Bekenstein-Hawking entropy scales with area, not volume. Remember that at its inception, it was characterized as the Shannon entropy of all possible configurations of matter collapsing into a black hole of fixed mass, angular momentum, and charge. Therefore, as an information measure over interior degrees of freedom, including those of the matter that formed the black hole and additional so-called gravitational degrees of freedom, Bekenstein-Hawking entropy enacts a lower-dimensional bound on entropy density (see Bekenstein 1975; Bekenstein 2001). But perhaps Bekenstein-Hawking entropy has nothing to do with the black hole interior. Maybe it's an information measure solely over horizon degrees of freedom that are causally accessible to the exterior (see Jacobson 1999; Rovelli 2019).

0.3 Unified Framework of Black Hole Paradoxes

More often than not, scholars neglect to acknowledge ongoing discussions within black hole thermodynamics and statistical mechanics when analyzing the black hole information loss paradox, including the interpretation of Bekenstein-Hawking entropy, which is far from settled. They either state assumptions without proper justification or omit a specific philosophical position altogether. I attribute the massive confusion and misguided publicity over the premises parameterizing the paradox in the selective exclusion of prior commitments. To remedy the dialectical gridlock, I have formulated an overarching, unified framework, which I call "Black Hole Paradoxes", that integrates the debates over information loss and black hole thermodynamics/statistical mechanics and taxonomizes the relevant 'camps' or philosophical positions.

The methodology of this dissertation consists of three stages of analysis, each corresponding to a chapter: 1) grounding black hole information loss in appropriate information-theoretic principles, 2) formalizing the information loss paradox, and 3)

assessing the ideal resolution of the paradox through the lens of Bekenstein-Hawking entropy. In the final stage of analysis, I am able to achieve the following three goals: 1) organize prospective solutions based on the interpretation of Bekenstein-Hawking entropy, 2) erect a pyramid of nested black hole paradoxes, and 3) propose alternative guiding principles to provisionally determine the status of information loss in quantum gravity that more aptly capture the relevant paradigmatic clash and align with the overarching framework.

In Chapter 1, I set up the mainstream narrative and define essential information-theoretic terms. I examine several responses to the mainstream narrative and evaluate their merit. I argue that contrary to received wisdom, black hole information loss is predominantly a puzzle about the increase in global von Neumann entropy (i.e., a proxy for external entanglement) alongside the decrease in maximal Boltzmann entropy (i.e., a proxy for counting degrees of freedom). Then in Chapter 2, I raise the stakes and promote the puzzle to a paradox. I showcase how black hole evaporation exasperates a pressure point in the relationship between the interior and event horizon. Either the evolution of the interior is decoupled from the shrinking event horizon, in which case it can store the growing entanglement. Or, the evolution of the interior is inextricably linked to the size and ultimately the presence of the event horizon, in which case it can't store the growing entanglement. Black hole information loss as I've formulated it tries to have its cake and eat it too, culminating in what I denote the "paradox of phantom entanglement": entanglement between Hawking radiation and phantom degrees of freedom of the former black hole interior.

We might be tempted to give up on black hole evaporation at this juncture. Such an attitude would disenfranchise the majority of responses within the debate without offering a positive solution for how to move forward. Why should the refinement of Hawking's calculation to explain that the final thermal state isn't truly global, or, alternatively, that it isn't precisely thermal, be considered conceptually equivalent to a wholesale rejection of black hole evaporation? Hawking's approximation is ill-suited to describe the black hole after it reaches Planck mass and so does not completely describe black hole evaporation. Anyone is free to come in and modify the Planck-scale physics or hone the formalism and reproduce the spirit of Hawking's result as a low-energy limit to rescue semi-classical gravity from undermining itself.

For many researchers, Hawking radiation is the missing ingredient unifying quantum mechanics, general relativity, and thermodynamics. Thermodynamics describes the aggregate behavior of large statistical systems, so black hole thermodynamics lights a path towards quantum gravitational statistical mechanics. And quantum gravity is the theory that will ultimately expose what black holes are made of. As I investigate in Chapter 3, Bekenstein-Hawking entropy – the mysterious quantity that quantifies how large of statistical systems black holes really are and is intimately linked to horizon area – is the key to unlock the solution to the paradox and provide much needed momentum to quantum gravity research.

I proceed to probe various proposals striving to resolve the paradox, such as stable and decaying remnants (like Planckian black holes transitioning to white holes, see

Rovelli 2019) and complementarity-based solutions (including ER=EPR and entanglement wedge reconstruction, see Almheiri et al. 2021), with the aim of analyzing how various interpretations of Bekenstein-Hawking entropy are accommodated or excluded within each proposal. Various attempts to evade the phenomenon of phantom entanglement impel us to take a stance on the identity of black holes – are they defined solely by their horizons? – thus informing the interpretation of Bekenstein-Hawking entropy. Remnant solutions say no and deny that Bekenstein-Hawking entropy is exhaustive. However, complementarity-based solutions say yes and affirm that Bekenstein-Hawking entropy is indeed the extent of a black hole's information.

In short, quibbling over the plausibility of proposed solutions is futile without first justifying the implicit interpretation of Bekenstein-Hawking entropy. Now, we need a litmus test to vet the overpopulated proposal space, which mandates reformulating the paradox of phantom entanglement with premises that are suitable for quantum gravitational contexts. It's imperative to establish premises that are not downright impossible to reconcile, though they are realistically difficult because black hole statistical mechanics stands at the cutting edge of quantum gravity research.

Solving foundational conundrums calls for renewed convergence between the philosophy and physics worlds. That's where 'information' saves the day. Rather than taking "it from bit" too far by professing that information replaces elementary units of the ontology, we can exploit the contexts in which information à la entropy allows us to abstract away from detailed structure to still learn about universal behaviors (see Batterman 2021). In other words, information holds the promise of confluence among the special sciences, like thermodynamics, and fundamental physics. The way forward is to commit to black holes as precariously "ordinary" thermodynamic systems (see Wallace 2020).

Since in terrestrial contexts, thermodynamic entropy can be recast as a statistical measure over a system's equilibrated constituents (compatible with its macrostate), it's plausible to infer that Bekenstein-Hawking entropy also warrants a statistical interpretation of a black hole's equilibrated constituents. This reasoning demands that if radiation emitted from burning coal eventually contains information about its microscopic structure, then radiation emitted from a black hole (in particular, the system carrying Bekenstein-Hawking entropy) should likewise contain information about its microscopic structure. But due to the curveball of trans-horizon entanglement, getting information to the right place inevitably requires novel and/or nonlocal physics (see e.g., Polchinski 2017).

At this stage in the project, rather than proving there's still a paradox, my goal is to proffer a formulation that explains where key assumptions diverge and accommodates prospective solutions – until we gain more insights about Bekenstein-Hawking degrees of freedom – without the need to deny any of the premises. Then whether the black hole information loss paradox upholds its status in quantum gravity is a venture for future investigation.

Chapter 1

Black Hole Information is Lost: No Matter?

1.1 Introduction: The Information Age

In the spirit of the information age, critics have accused the black hole information loss paradox of being a meme that has gone rogue. The metaphorical tweet that has gone viral reads something like this: Information about anything unfortunate enough to find itself inside a black hole is forever lost to the outside world when the black hole radiates away and disappears. Right, makes sense. But so what? Black holes are supposed to be cosmic one-way streets.

What's lacking in the literature is a rigorous *conceptual* articulation of why black hole information loss is considered by some to be paradoxical. Usually, the problem is stated in mathematical terms as a breach of global unitarity, or a pure-to-mixed transition, without a clear demonstration of how the underlying framework self-destructs. Wallace (2020) observes that unitarity is sometimes taken to be a pillar of quantum mechanics; therefore, rejecting unitarity is apparently tantamount to rejecting quantum mechanics. This is the argument grounding non-unitary evolution from pre-to-post-evaporation, what he labels the "evaporation-time paradox" (Wallace, 2020, p. 220). However, given the option of simply expanding the scope of quantum mechanics, he's not swayed by this reasoning. He avers,

> Information loss seems prima facie plausible, and in any case the question seems to require a full understanding of quantum gravity to answer and so may be premature (Wallace, 2020, p. 210).

Wallace (2020) is right in calling out an aversion to non-unitary as a lazy formulation of the black hole information loss paradox. The mainstream narrative of information loss, which Hawking (1976) initiated and others perpetuated, concerns a breakdown of predictability and retrodictability. Though the observation about indeterminism is not wrong, surely labeling it paradoxical can't have bite beyond individual metaphysical preferences about the ideal form of laws of nature. In fact, Okon

and Sudarsky (2017) protest that the mainstream narrative is biased towards quantum theories that hold on to unitarity, like the Many Worlds Interpretation (MWI), despite the fact that the measurement problem hasn't been solved.

Many philosophers of physics have also scoffed at the appropriation of the term 'information' to frame the purported paradox. Belot et al. (1999), for example, are far from persuaded that information has anything to do with black hole evaporation. If anything, 'information loss' has made for a viral – or brilliant – PR campaign, however you want to view it.

> Speaking loosely, we might say that information about the universe is lost in the course of black hole evaporation, and join the vulgar in labelling Hawking's result 'the Hawking Information Loss Paradox' (Belot et al., 1999, p. 190).

I myself have been tempted to echo their sentiment in calling the terminology vulgar based on haphazard usage in the literature. However, by dissecting a formal definition of unitarity, I've identified relevant, information-theoretic implications of black hole evaporation. I've come to realize that information theory is exactly what the discourse needs for that sought-after, rigorous conceptual articulation of a paradox.

To assuage lingering skepticism towards post-evaporation information loss being puzzling in the first place, I have undertaken a more detailed examination of black hole evaporation confined to semi-classical gravity. My objective for this paper is to fill in the gaps of how information, when accurately defined, unveils genuinely unpalatable consequences that sow the seeds of a potent, black hole information loss paradox. I contend that indeterminism is misconstrued as the root of the puzzle when it's merely an unintended side effect of a more pernicious phenomenon: matter that goes missing after the black hole evaporates and left-over radiation that apparently never gets the memo.

My argument unfolds as follows. In Section 1.2, I present a sweeping overview of the methodological commitments necessary to frame the evaporation-time puzzle and unearth the impending pressure points. I then expeditiously walk through Hawking's derivation of black hole evaporation in Section 1.3. Subsequently, in Section 1.4, I put forth original arguments demonstrating how unitarity incorporates four distinct principles of information conservation: 1) constancy of coarse-graining, 2) deterministic evolution, 3) conservation of degrees of freedom, and 4) preservation of entanglement. The black hole information puzzle is about the distastefulness of non-unitarity after all, so it's necessary to isolate which component of unitarity is being violated.

Section 1.5 contains the most substantive portion of my contribution. Based on the framework I develop, I assess whether black hole evaporation violates each of the four information conservation principles in the following ways: 1) thermodynamic evolution, 2) indeterministic dynamics, 3) elimination of degrees of freedom, and 4) appearance of external entanglement. I discover that black hole evaporation violates the first two of these principles as downstream effects of violating the third and fourth of these principles, the levels which have the most explanatory power.

I contrast my analysis with that of the mainstream narrative, whose focus is the violation of the second principle. I show that by singling out determinism as the only relevant component, the mainstream narrative runs into a sticky contradiction that no amount of poetic quotes by Hawking himself can gloss over. I then argue that the root of the puzzle pertains to spontaneous appearance of global entanglement. A violation of unitarity at this deeper level signals that the underlying ontology is unstable, and I argue that it legitimizes investigating an evaporation-time paradox within the domain of semi-classical gravity.

1.2 Methodological Commitments: Quantum States and Unitarity

In order to begin engaging with the majority of the black hole information loss discourse charitably, we must pay attention to the perceived problem: the loss of something important. Solving the problem entails bringing about the opposite: the conservation of something important. Whatever that important thing is, we need to be able to keep track of it.

Physics has a time-tested way of keeping track of important things: by identifying a target system and specifying its state. A state relays information about the target system's configuration. It does so by relating its parts occupying different spatial locations into a *unified whole at an instant of time*. As Curiel (2021) puts it,

> A state, therefore, can be thought of as a set of the values of quantities that jointly suffice for the identification of the species of the system and for its individuation at a moment. As such, the state is the most fundamental unit of theoretical representation of a system as a unified system, rather than just as (say) a bunch of random, unrelated properties associated with a spatiotemporal region (p. 3).

States record the values of degrees of freedom instantaneously, or to be mathematically careful, over an infinitesimal period of time.[1] Two elements of this definition demand unpacking: 'degree of freedom' and 'instantaneous'. First, a degree of freedom is an ontological and modal concept. It's ontological because it quantifies objects and properties in the *actual* world, like a snapshot. It's also modal because 'freedom' refers to the *possible* behaviors of these material entities, i.e., the range of values that free variables can take on.[2] In quantum field theory, for instance, particle number and mode/frequency in bounded or unbounded spatial regions are degrees of freedom in the ontological sense as physical observables, as well as in the modal sense in that each possible value represents a degree of freedom in and of itself.

[1] Albert (2000), for instance, prefers to replace states with infinitesimal dynamical conditions.

[2] Though a conceptual analysis of 'degree of freedom' deserves far more attention, I'd like to thank Baptiste Le Bihan for inspiration in getting started.

Second, one might worry that the dependence of state determination on simultaneity poses a challenge to the relativity of simultaneity. However, the metrical definition of zero temporal distance is tied to a particular coordinate system, so a foliation into states must also be tied to a particular coordinate system. Since relativity theory doesn't privilege any foliation, the universe permits a multiplicity of foliations, each with its own set of states.

Moreover, the worry is unfounded so as long as the dynamical laws retain the same form under coordinate transformations. In fact, to deduce what is or isn't conserved over time, we must heed the dynamical laws' symmetries, which maintain the laws' mathematical form under systematic changes to state values. Examples of such systematic changes include linear translations in time, linear and angular translations in space, and velocity/acceleration boosts.

Now, we need a formalism for state specification and dynamical laws that allow us to compare states. After all, we're trying to keep track of what is or isn't conserved, which calls for a more in-depth understanding of temporal evolution. In a semi-classical framework, both of these components are largely dictated by relativistic quantum field theory, so let's walk through the formalism to gauge our inherited methodological commitments.[3]

A system is represented by a complex Hilbert space $\mathcal{H}$, where the inner product of vectors defines relative distance. The dimensionality of $\mathcal{H}$ is equal to the cardinality of the orthonormal basis, which is constructed from a set of n linearly independent vectors ψ_i that span the space. These orthonormal basis vectors are eigenstates of Hermitian operators associated with chosen physical observables and represent independent degrees of freedom. All other vectors are linear combinations, such as superpositions, of the basis vectors.

I should interrupt with a disclaimer. Quantum fields technically have infinite-dimensional, non-separable Hilbert spaces because any continuous spatial region, bounded or unbounded, contains uncountably infinite degrees of freedom. But, non-separable Hilbert spaces pose practical and philosophical issues for adequate physical representation. So, a strategy to impose separability is to discretize the eigenvalue spectra of the observables generating the eigenvectors (see Ruetsche 2011).

Fock space, which is used in Hawking's derivation, is a type of Hilbert space with discrete eigenstates of the joint observable of particle number and frequency (a proxy for energy). It's a special construction that combines many single-particle Hilbert spaces to admit of superpositions of total particle number/energy, where the physical picture of a classical field is replaced with that of indeterminate quantum oscillators. What's noteworthy is that a system with countably infinite degrees of freedom taking up infinite volume has only finite density; so now, a bounded continuous spatial region contains only finite degrees of freedom.

Considering that it's much easier to employ the resources of standard quantum theory for systems of finite density, we can proceed with characterizing states. They

[3]For an introductory treatment of the mathematical formalism of quantum theory, see Albert (1992). For a more technical treatment, see Ruetsche (2011).

can have one of two mathematical properties: purity or mixedness. Pure states are associated with unit-length vectors representing microstates; they encode the value of a wavefunction, Ψ, through their location in $\mathcal{H}$ (see Equation 1.1). The coordinates are given by the complex coefficients or amplitudes, α_i, of the orthonormal basis vectors in the spectral decomposition:

$$|\Psi\rangle = \sum_{i=1}^{n} \alpha_i \psi_i. \tag{1.1}$$

The norm of the vector is $\|\Psi\| = \sum_{i=1}^{n} |\alpha_i|^2$ and is usually normalized to one so that the square of the amplitudes can be interpreted as real-numbered probabilities of measurement outcomes (ie., Born's rule). Thus far, the most rational designation for states of the global system, i.e., the entire universe, is pure, a widely accepted convention (see e.g., Page 1995; Jaksland 2021). The entire universe is represented by unit vectors in an overarching Hilbert space $\mathcal{H}$, which cannot be embedded into another Hilbert space without imposing unphysical degrees of freedom. If $\mathcal{H}$ is factorizable into a tensor product of Hilbert subspaces H_i, in which a subspace represents the physical observables of a subsystem, then it also embeds the possible pure states of the subsystems (see Equation 1.2):

$$\mathcal{H} = H_1 \otimes H_2 \otimes \cdots \otimes H_i. \tag{1.2}$$

The surface that forms by rotating a unit vector is the boundary of a unit-radius Bloch hypersphere – the quantum analogue of an energy hypersurface in classical phase space. Pure states always intersect the surface of the hypersphere, but mixed states, on the other hand, are associated with vectors in the interior that are less than unit length. The magnitudes of these vectors help reconstruct the mixed states' density matrices, ρ, that encode weighted distributions over collections of pure states (see Equation 1.3). The sum of the coefficients, p_i, is also normalized to one:

$$\rho = \sum_{i=1}^{n} p_i \, |\psi_i\rangle \, \langle\psi_i| \, ; \; p_i \geq 0, \sum_{i=1}^{n} p_i = 1. \tag{1.3}$$

The physical interpretation of the weighted distribution is context-dependent. When the distribution is sharply peaked, with one of the vectors weighted by $p = 1$, the density matrix is equivalent to that vector and describes a pure state. A density matrix describes a mixed state only when two or more vectors in the collection are positively weighted.

A density matrix is then the result of a projection onto a Hilbert space H, usually a subspace of fewer desired observables than $\mathcal{H}$, with fewer linearly independent eigenvectors acting as independent degrees of freedom, and therefore, of lower dimensionality. Whereas vectors correspond to one-dimensional Hilbert subspaces, density matrices fittingly accommodate higher-dimensional Hilbert subspaces. Indeed, $\mathcal{H}$ can be decomposed into a direct sum of subspaces H_i obtained through projection (see Equation 1.4):

$$\mathcal{H} = H_1 \oplus H_2 \oplus \cdots \oplus H_i. \tag{1.4}$$

Even though these subspaces technically have different dimensionalities, their associated density matrices are made to have the same number of coordinates as $\mathcal{H}$ by including ψ_i's for which $p_i = 0$. When $p_i = 0$, the coordinate of an unwanted degree of freedom is being set to zero.

Mixed states are underdetermined in their physical interpretation. One application is for statistical ensembles of microstates belonging to macrostates, in which the microstates share certain macroscopic parameters like particle number and energy. According to those inclined towards an epistemic view of macrostates, the probability distribution quantifies our uncertainty of the system's pure state, each vector being a weighted possibility. The interpretation of a macrostate is readily apparent when the collection of vectors in a density matrix is a proper mixture, i.e., a subset of the orthonormal basis vectors of the overarching Hilbert space, $\mathcal{H}$. In fact, Bloch vectors can only represent mixed states that are proper mixtures.

However, when the collection of vectors in a density matrix is not a subset of $\mathcal{H}$'s basis vectors, i.e., the eigenvectors of the chosen observables, it's an improper mixture and reflects the reduced density matrix of a subspace H_i associated with the states of a subsystem. In this scenario, the rest of $\mathcal{H}$ representing other subsystems has been intentionally neglected or 'traced out'. For such mixed states, the reduced density matrix does not correspond to a subspace in a genuine tensor product building up $\mathcal{H}$. The physical interpretation here is that the subsystem being described by an improper mixture contains external correlations and the weighted distribution now quantifies the extent of quantum entanglement with the traced out subsystems.

Entanglement is the cornerstone of black hole evaporation, so let's take a brief detour into its mathematical and physical interpretations. Entanglement is a mathematical property distinctive of pure states. It designates a holistic constraint, where the individual values of proper subsets of degrees of freedom supervene on the joint values of the complete set of degrees of freedom. There's admittedly a confusing vagueness surrounding its physical interpretation and whether the adjective 'entangled' refers to systems with internal or external entanglement. Entanglement encapsulates nonlocal correlations, and since 'relation' is part and parcel of 'correlation', I take entanglement to be a physical relation fixing the nonlocal correlation structure among the properties of its related entities.

That being said, the holistic constraint inherent to entanglement has major ramifications for the state specification of entangled subsystems. The best we can do is invoke reduced density matrices to report the values of proper subsets of degrees of freedom and the extent of entanglement with traced out subsystems. But we forego the nonlocal correlation structure because the mixed state of an entangled subsystem is parasitic on the pure state of the total system. That's why it's impossible to distinguish entangled subsystems from the macrostates of non-entangled subsystems just by looking at the density matrix of a mixed state (see e.g., Polchinski 2017). Indeed, there's a special class of mixed states where the subsystems of interest are entangled and in thermodynamic macrostates. These Goldilocks states are exactly thermal and of paramount importance to black hole evaporation.

The payoff of all this machinery is its behavior under unitarity, the set of mathematical operations in Hilbert space generating the dynamics and symmetries. Unitary operators time-evolve quantum states in obedience with the Schrödinger Equation. The formal definition of a unitary operator, U, has the following four requirements.

Preconditions of Unitarity:

1. Linearity: Allows unitary operators to act on superpositions of states, just as they act on the individual components and their amplitudes;

2. Boundedness: Guarantees that unitary operators implement continuous transformations, and the objects they act upon stay in the same Hilbert space;

3. Surjectivity: Establishes that unitary operators uniquely map input states to output states;

4. Norm-preservation: Ensures that unitary operators satisfy the axioms of probability.

What follows from this definition is that the inverse of a unitary time-evolution operator, U^{-1}, which generates time-reversed evolution, is itself unitary. The symmetries ensure that pure states stay pure and mixed states stay mixed. The revelation of the conservation of state properties is the key to understanding black hole information loss and will be the launching platform for Section 1.4.

1.3 Evaporation-Time Puzzle

With the methodological commitments laid bare, I can now construct the scaffolding of the evaporation-time puzzle. It will become synonymous with the black hole information loss puzzle by the end of Section 1.5 after I've motivated information-theoretic arguments. I'm also purposefully labeling it a puzzle at this stage and refraining from justifying its paradoxical status until Chapter 2.

The purpose of this section is to acquaint the unfamiliar reader with black hole evaporation. The exposition is mostly qualitative, and while I don't presuppose a rigorous physics background on the part of the reader, familiarity with the technical jargon and theory-specific vocabulary of quantum mechanics, general relativity, thermodynamics, and statistical mechanics is highly beneficial.

1.3.1 Black Hole Evaporation in Hawking's Semi-Classical Framework

In his inaugural paper on black hole evaporation, Hawking (1975) demonstrates a remarkable transmutability between matter and spacetime. Quantum matter fields on a black hole spacetime extract energy from the underlying geometry for particle

production. Once matter has collapsed to form and become trapped inside an event horizon, this process involves the unprecedented transition of the exterior region from a vacuum state (absence of particles) to a radiation state (presence of particles) at a characteristic temperature due to the intervening curvature induced by the black hole. It's as if the vacuum fields split at the event horizon into positive-frequency modes propagating towards future infinity that are entangled with negative-frequency modes approaching the black hole singularity.

The derivation of Hawking radiation only requires quantum field theory on curved spacetime, in which the matter fields are essentially skating on geodesics, the curved analogue of straight lines, but uncoupled to the metric. We cannot yet infer a transfer of energy from the metric to the matter fields in the production of Hawking radiation. Therefore, fortifying the prediction to black hole evaporation as a reciprocal effect, in which the reduction of a black hole's mass by the absorption of negative energy exactly compensates for the emission of positive energy, brings us into the domain of semi-classical gravity incorporating the back-reaction of spacetime. Although Hawking himself did not utilize the semi-classical Einstein field equation (SEFE) (refer to Equation 1.5), SEFE calculations have subsequently been carried out (see Wallace 2018). The classical Einstein tensor, $G_{\mu\nu}$, is coupled to the expectation value of an energy-momentum operator defined on the matter fields $\left\langle \hat{T}_{\mu\nu} \right\rangle_{\psi}$:

$$G_{\mu\nu} = \frac{8\pi G}{c^4} \left\langle \hat{T}_{\mu\nu} \right\rangle_{\psi}. \tag{1.5}$$

Keep in mind though that back-reaction effects must be adequately low-energy, and Hawking (1975) is upfront about the trustworthiness of his calculation. When the evaporating black hole shrinks down to Planck mass, about 10^{-5} g, the semi-classical approximation is no longer valid. Nonetheless, Hawking stipulates that there's not much else for the black hole to do except disappear entirely. Even if there were no singularity that extinguished the collapsing matter, the remaining mass is insufficient to resurrect it. Yes, Hawking is stretching the scope of semi-classical gravity here to argue that black holes evaporate completely and leave behind no remnants. Yet until we have compelling reasons to believe otherwise, we're taking his word at face value based on epistemic authority.

For a Schwarzschild black hole, the temperature of Hawking radiation T_H at future infinity is proportional to its surface gravity κ. The first law of black hole mechanics, which relates quasi-static changes in mass dM to changes in surface area dA, $dM = \frac{\kappa}{8\pi} dA$, looks remarkably like Clausius's theorem, which relates quasi-static changes in energy dE to changes in entropy dS, $dE = T_H dS$. From this parallel and mass-energy equivalence, Hawking deduced that a black hole's dimensionless entropy scales with the magnitude of its surface area (refer to Equation 1.6):

$$S_{BH} = \frac{c^3 A}{4G\hbar}. \tag{1.6}$$

He thereby bolstered Bekenstein's intuition that black holes have entropy, fixed the proportionality constant, and the physics community christened the quantity as

Bekenstein-Hawking entropy. Not only does Hawking temperature lend credibility to the idea of Bekenstein-Hawking entropy being a thermodynamic property, the amalgamation of fundamental constants (speed of light c, gravitational constant G, and reduced Planck constant $\hbar$) provocatively intimates that Bekenstein-Hawking entropy is also a statistical measure over a finite number of quantum gravitational degrees of freedom (see Carlip 2014). I will delve more into black hole thermodynamics and statistical mechanics in Chapter 3.

A puzzle arises (and I'm purposefully labeling it a puzzle at this stage, not a paradox) because the causal barrier posed by the event horizon implies that black hole evaporation is an irreversible process. The collapsing matter and negative-energy modes that got trapped inside the black hole can't be converted into radiation since their required escape velocity would exceed the speed of light. Note that classically, black holes have no hair. What this idea of no hair means is that after a generic black hole evaporates, the only information left in Hawking radiation encodes its macroscopic properties of mass, angular momentum, and charge, all of which depend on the radius of the event horizon and no other fine-grained details. Black holes of the same macroscopic parameters have statistically indistinguishable Hawking radiation that's insensitive to the details of the gravitational source, such as the composition of the collapsing matter, as well as the details of how the vacuum state at the event horizon breaks apart into positive and negative-frequency modes.

1.3.2 Pure-to-Mixed Evolution

The phrase that has been thrown around to indicate information loss in black hole evaporation is 'non-unitary' or 'pure-to-mixed' evolution. The initial black hole and surrounding vacuum are jointly a pure state, yet the final radiation state is a mixed state. As it turns out, the final radiation state is not just mixed; it looks exactly thermal.[4] The tricky feature about thermal states is that we can't rule out potential entanglement. Without further insights about the overarching Hilbert space and how it factorizes, it's unclear whether what looks to be thermal Hawking radiation is in an equilibrium macrostate, highly entangled, or both.

I will not present the density matrix that Hawking (1976) calculated due to the breadth required, but here's a summary of its implications. The final radiation state can be interpreted as a higher-level, non-uniform probability distribution over energy macrostates of various radiation subsystems emitted at different frequencies and particle numbers, holding fixed the overall temperature that's inversely proportional to the black hole's initial mass. The microstates compatible with the energy macrostate of each subsystem, however, are equally likely. Seeing as the positive-frequency modes are what's left after the black hole evaporates in finite time (the negative-frequency modes having been trapped behind the event horizon), it seems reasonable to conclude that the thermal radiation at future infinity comprises the global state.

[4]For simplicity's sake, I'm neglecting gray-body factors that deviate from thermality due to mode-dependent tunneling rates across the effective potential barrier (see Hawking 1976; Wallace 2018).

There's a deep mystery about why final radiation states are mixed and not pure, as global microstates are expected to be. It's impossible to tell unambiguously just by looking at a mixed state whether it's describing an entangled subsystem or a self-contained system in an unknown pure state. Either way, mixed states incur an information deficit as compared to pure states. Hawking (1976) reflects,

> Because part of the information about the state of the system is lost down the hole, the final situation is represented by a density matrix rather than a pure quantum state (Hawking, 1976, p. 2460).

Though interpreting mixed states in isolation is futile, situating them in a dynamical context could help narrow down the physical interpretation. Back in Section 1.2, I asserted that to get to the bottom of the evaporation-time puzzle, we must ascertain the loss of something important by comparing states. Hawking himself reveals an important clue – information is lost because it went "down the hole".

1.4 Information Conservation Principles

Notice how information loss is intertwined with irreversibility. The explanation in the literature for this has predominantly been cashed out as a breakdown of determinism (see Belot et al. 1999; Manchak and Weatherall 2018; Maudlin 2017; Wallace 2020). Hawking himself stuck to that story even 40 years after first telling it.

> Forty years ago, one of the authors argued [1] that information is destroyed when a black hole is formed and subsequently evaporates [2, 3]. This conclusion seems to follow inescapably from an 'unquestionable' set of general assumptions such as causality, the uncertainty principle and the equivalence principle. However it leaves us bereft of deterministic laws to describe the universe. This is the infamous information paradox (Hawking et al., 2016, p. 1079)

Maudlin (2017) chimes in affirmatively:

> Despite the usual moniker, the paradox is only tangentially about information, and the analytical tools of information theory (Shannon or otherwise) are not relevant to it. A much better characterization of the problem adverts to an apparent breakdown of either determinism or of unitary evolution of the quantum state (p. 2).

Admittedly, what remains ambiguous in the puzzle is how the concept of information is being employed – the loss of which is supposed to be paradoxical. However, I endeavor to show in Section 1.5 that the prevalent account of indeterminism is descriptive, not explanatory. My goal in this section is to fill in the gaps about how information, properly construed, plays a significant. multifaceted role in the black hole information puzzle.

I left off in Section 1.2 picking up on the fact that unitarity guarantees the purity and mixedness of states. This seemingly simple and straightforward statement is the foundation for the relationship between unitarity and information conservation. I propose four plausible connections: 1) constancy of coarse-graining, 2) deterministic evolution, 3) conservation of degrees of freedom, and 4) preservation of entanglement.

Constancy of Coarse-Graining: Unitarity is the quantum analogue of Liouville's theorem, which maintains the constancy of coarse-graining by conserving the Boltzmann entropy of an ensemble of microstates compatible with a fixed set of observables. Classically, microstates are points in a high-dimensional phase space, which assigns a degree of freedom to each spatial component of position and momentum per particle. Constraining some of the system's parameters that depend on position and momentum, such as energy, carves out a hypersurface in phase space, whose volume serves as a proxy to count all physical possibilities associated with that macrostate. Boltzmann entropy is simply the logarithm of any hypersurface's phase space volume, at least if its scope is expanded beyond thermodynamic applications.

Classical phase space volume is representationally similar to Hilbert space dimensionality, in that both cluster microstates into macrostates (see Sheldon Goldstein (2020)). So in a Hilbert space construction, whether it's the overarching space $\mathcal{H}$ or a subspace H_i, Boltzmann entropy S_B is defined as the logarithm of its dimensionality (refer to Equation 1.7):

$$S_B = \ln(dim H). \tag{1.7}$$

Like in the classical case, the dimensionality of the state space quantifies the number of degrees of freedom. But classical microstates are distinct from phase space degrees of freedom, whereas they coincide with Hilbert space degrees of freedom, indicating a classical-to-quantum transition in the interpretation of Boltzmann entropy. The reason classical microstates transform into quantum degrees of freedom is because they're eigenevectors of definite measurement outcomes. All other exclusively quantum microstates are superpositions of these definite measurement outcomes, and therefore, not independent degrees of freedom. Although there has been recent interest in formalizing quantum macrostates (see e.g., Teufel 2022), the strategy that's of relevance to the black hole information loss discourse is retaining classical macrostates by grouping together eigenvectors of relevant observables (such as horizon area, mass/energy, temperature, etc.) and projecting onto Hilbert "macrospaces" (see Sheldon Goldstein 2020).

Circling back to the analogy between Liouville's theorem and unitarity, they both establish that Boltzmann entropy is continuously held constant under time-reversal invariant dynamics. If we're tracking the dynamical trajectory starting from a unique microstate, i.e., a point in phase space, time-symmetric laws ensure that we always end up with another unique microstate associated with another point in phase space. If we're tracking an ensemble of trajectories starting from an initial macrostate, we always end up with an ensemble of microstates occupying the same volume as the initial

macrostate. Similarly, unitarity performs the role of Liouville's theorem for Hilbert spaces. After projecting onto a Hilbert macrospace, unitarily evolving it doesn't disturb its dimensionality or Boltzmann entropy. That's because acting with a unitary operator on a density matrix of n nonzero entries, corresponding to an initial microstate for $n = 1$ or initial macrostate for $n \geq 2$, will produce another Hilbert macrospace of the same dimensionsality containing the evolved state vector(s) of the original space (see Equation 1.8):

$$U : \rho_j \rightarrow \rho_k;\ dimH_j = dimH_k \geq 1. \tag{1.8}$$

Alarms may be ringing in your mind that Liouville's theorem and unitarity are at odds with the Second Law of Thermodynamics. You would not be wrong. One of the great challenges of physics has been to derive a time-directed emergent law from time-reversal invariant fundamental laws (see Albert 2000; Wallace 2023). It may come as a relief then that the constancy of coarse-graining doesn't deliver physically meaningful macrostates throughout the universe's evolution.

The initial macrostate is often identified by specifying useful, high-level properties, notably thermodynamic constraints like fixed energy or temperature. Yet as time goes on, individual dynamical trajectories appear fibrillated in classical phase space and no longer share properties exclusive to their ensemble. In Hilbert space, the original state vectors evolve into a subspace with eigenvectors of a contrived and bizarre composite observable.

An explanatorily better strategy is to coarse-grain the state space not just to demarcate the initial conditions but also to demarcate macrostates throughout the evolution. By imposing constraints on acceptable collective parameters at later times, motivated by theory, practicality, or both, coarse-graining becomes variable. Microstates in the original ensemble end up being regrouped into different macrostates at later times, most likely those with higher or maximal entropy indicative of equilibration and equilibrium. Consequently, a benign violation of unitarity occurs at the level of effective macrodynamics.

To reconcile emergent time asymmetries with fundamental time symmetries, and moreover, to accurately encapsulate the pervasiveness of thermodynamic phenomena, probability takes center stage, leading us down the path to information theory. Recall from Section 1.2 that the quantum formalism already utilizes density matrices as mixed states to assign probability distributions to pure states, embedding probability as epistemic uncertainty due to ignorance. Therefore, by linking ensemble size to an uncertainty measure over microstates, an information-theoretic interpretation of macrostates arises organically from the very definition of a mixed state. Shannon (1948) asks,

> Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome? (p. 389).

That measure takes the form of Shannon entropy, and its statistical mechanical incarnation with the inclusion of Boltzmann's constant k_B is known as Gibbs entropy

S_G:

$$S_G = -k_B \sum_{i=1}^{n} p_i \ln p_i;\ p_i \geq 0,\ \Sigma_i p_i = 1. \tag{1.9}$$

Pure states with maximally concentrated probability distributions (i.e., exact microstates) have zero Gibbs entropy and mixed states with spread out probability distributions (i.e., macrostates) have positive Gibbs entropy.

In fact, I take Gibbs entropy to add nuance to Boltzmann entropy for discerning a system's thermodynamic behavior. For example, a maximally mixed state with a uniform probability distribution is in the microcanonical ensemble, signaling an equilibrium macrostate at fixed energy. A stable energy eigenstate is indicative of a system's effective isolation, and if the system satisfies the properties of internal thermodynamic equilibrium, like balanced occupation of accessible modes among subsystems, then each configuration is equally likely.[5] Maximally-mixed states actually saturate their Gibbs entropy and recover Boltzmann entropy (refer to Equation 1.10):

$$p = \frac{1}{n} \rightarrow S_G = S_B = \ln(n). \tag{1.10}$$

A thermal state, however, incorporates the non-uniform Boltzmann probability distribution and belongs to the canonical ensemble, signaling an equilibrium macrostate at fixed temperature and average energy. Fixed temperature T is indicative of a system's coupling with its environment, often idealized as a vast external reservoir or heat bath. As opposed to the microcanonical ensemble, thermal systems can access a range of energy eigenstates E_i, but higher-frequency modes are suppressed, as is characteristic of a thermal spectrum (refer to Equation 1.11):[6]

$$p_i = \frac{e^{-E_i/T}}{Z}. \tag{1.11}$$

Brillouin (2013) proposed identifying changes in Gibbs entropy with changes in available information, quipping that information is "negentropy" (p. 153). On the flip side, entropy is negative information, or better yet, I prefer to cast it as an information deficit. In other words, entropy captures the amount of information contained in a system, or its internal information content, that cannot be gleaned from information about its host ensemble. Brillouin's famous equation captures the inverse relationship between internal information I_{int} and Gibbs entropy S_G:

$$\Delta I_{int} = -\Delta S_G. \tag{1.12}$$

Susskind and Lindesay (2004) claim that unitarity implies a principle of information conservation, although they don't motivate why. The missing puzzle piece is that

[5] A quantum field in a stable energy eigenstate admits of a degeneracy of configurations involving different occupation numbers, i.e., different combinations of particle number per mode/frequency (see Ruetsche 2011).

[6] Z is the partition function that normalizes the probabilities.

over and above Boltzmann entropy, unitarity also continuously conserves Gibbs entropy. Therefore, the link follows from an information-theoretic interpretation of Gibbs entropy and the fact that unitarity constrains both ΔS_G and ΔI_{int} to remain zero. This result is powerful when applied to the universe as a whole. Whereas unitarity can be broken for temporarily isolated subsystems through external interactions, the universe is permanently self-contained and bound by unitarity as well as the principle of information conservation.

> **Global Unitarity (GU)**: The Gibbs entropy of the state space associated with the universe's initial micro-/macrostate remains constant at every infinitesimal time interval.

To reiterate, GU is consistent with the Second Law when it involves benign violations at the level of effective macrodynamics due to increasing uncertainty. Going even further, unitarity utterly fails to provide predictive power to ascertain the macroparameters and entropy of the final macrostate even when we're provided the macroparameters and entropy of the initial macrostate. Rather, we need to apply a probability distribution over histories instead of microstates to induce branching behavior towards higher-entropy and equilibrium macrostates (Wallace, 2023), insinuating that information is not conserved because $\Delta S > 0$ and $\Delta I < 0$.

At this juncture, I want to expose two metaphysical assumptions concerning the nature of dynamical laws and the quantification of fundamental ontology that are embedded in unitarity. A nontrivial violation of unitarity requires rejecting one or both of these components, which is why it's prudent to decouple them as separate conditions for information conservation.

Deterministic Dynamics: Unitarity generates deterministic evolution, as is apparent from the requisites of surjectivity and norm-preservation. In the Schrödinger picture, unitary operators smoothly rotate vectors in Hilbert space (more precisely, the Bloch hypersphere), securing a one-to-one mapping between input and output states for any time interval. Taking the example of generic vectors associated with density matrices, smooth rotation from time t_0 to time t continuously evolves amplitudes without affecting the length $\|\rho(t)\|_1$ (refer to Equation 1.13):

$$U(t) : \ \rho_0 \to \rho(t); \ \frac{d}{dt}\|\rho(t)\|_1 = 0. \tag{1.13}$$

Thus, inverse-time evolution operators are also unitary that smoothly reverse-rotate vectors, allowing for the exact amplitudes of any state to be recovered (refer to Equation 1.14):

$$U^{-1}(t) : \ \rho(t) \to \rho_0; \ \frac{d}{dt}\|\rho(t)\|_1 = 0. \tag{1.14}$$

It's clear that physicists and philosophers contemplating black hole information loss identify unitarity with deterministic evolution. Susskind and Lindesay (2004) emphasize predictability and retrodictability:

> One way of stating the principle of information conservation is through the unitarity of the S-matrix. The point is that a unitary matrix has an inverse, so that in principle the initial state can be recovered from the final state (pp. 81-82).

Susskind and Lindesay (2004) intimate that deterministic evolution is the mechanism by which information is conserved. Maudlin (2017) elucidates this claim:

> There is an obvious sense in which any dynamical evolution that is deterministic in both time directions 'preserves information'. In such a case, the value of the state at any time implies the value of the state at any other time (p. 3).

Since state specification encodes the values of all degrees of freedom, and deterministic laws uniquely map one state onto another (the output of the mapping depending on the chosen time interval), then it must be the case that only the values of the degrees of freedom change, not the quantity of degrees of freedom. For the universal wavefunction to be subject to determinism, the dynamical laws must be expressed as unitary time-evolution operators, and the global quantum state must be pure.

> **Global Unitary Evolution (GUE)**: For any arbitrarily chosen time interval, the dynamical laws take the form of a one-to-one map: 1) time-evolution operators uniquely produce output global state vectors from input global state vectors, and 2) inverse time-evolution operators uniquely recover input global state vectors from output global state vectors.

Of course, the most obvious strategy to violate GUE is to reject unitary time-evolution operators, as seen in measurement or spontaneous collapse approaches to quantum theory. Indeterministic dynamical laws can take the form of one-to-many maps, resulting in the loss of predictability, as well as many-to-one maps, resulting in the loss of retrodictability. Now, the Boltzmann entropy of the state space associated with the universe's initial ensemble no longer remains constant at every infinitesimal time interval. One-to-many maps cause an expansion of the initial ensemble, and therefore, an increase in Boltzmann entropy, whereas many-to-one maps cause a contraction of the initial ensemble, and therefore, a decrease in Boltzmann entropy. These changes in Boltzmann entropy also affect Gibbs entropy. Consequently, the constancy of coarse-is violated not just at the level of macrodynamics recovering the Second Law but also at the level of microdynamics, in which $\Delta I_{int} = -\Delta S_G \neq 0$.

However, there's another more subtle and arguably more insidious strategy to violate GUE. Notice how it was taken for granted that once we construct the Hilbert space of a closed system, specify an initial state, and let it evolve, its dynamical trajectory stays in that space. The stability of the overarching Hilbert space is another, more basic assumption, and reflects a deep metaphysical principle with historical precedent.

Conservation of Degrees of Freedom: Unitarity, at its foundation, protects the long-standing metaphysical principle that the fundamental ontology is neither created nor destroyed. That is to say, degrees of freedom (or more concretely and without making any claims about the fundamental ontology, the physical bearers of degrees of freedom) are neither created nor destroyed. For a choice of basis, the requisite of boundedness confines a system's dynamics to a unique overarching Hilbert space $\mathcal{H}$.

The quantity of information required for exhaustive state specification reflects the maximum number of vector coordinates, which is a straightforward reflection of $\mathcal{H}$'s dimensionality and Boltzmann entropy. Susskind and Lindesay (2004) illustriously conceptualize a system's maximal Boltzmann entropy as its information storage capacity. Pure states maximize Boltzmann entropy and cap information storage capacity because they need as many coordinates for their representation as the dimensionality of $\mathcal{H}$ in which they live, so they're assembled out of the maximum number of degrees of freedom.

Mixed states, in contrast, have a more complicated relationship with Boltzmann entropy. Statistical ensembles usually have less than maximal Boltzmann entropy in order for the delineated macrostates to be useful and vet at least some unwanted degrees of freedom. One way to vet unwanted degrees of freedom is to set aside certain eigenstates of the overarching Hilbert space such that the remaining collection is a proper mixture and indicates a global macrostate. Another way is to trace them out, in which the statistical ensemble would also be an improper mixture and indicate the macrostate of a subsystem. For entangled subsystems, Boltzmann entropy is always less than maximal because by definition, subsystems contain a proper subset of degrees of freedom. It's inferred by projecting onto the lowest-dimensional Hilbert subspace spanned by all the positively weighted vectors forming an improper mixture.

Changes in information storage capacity reflect inserted or discarded degrees of freedom, which are positively correlated with changes in Boltzmann entropy. I will denote information storage capacity as C to distinguish it from I_{int} as negentropy:

$$\Delta C = \Delta S_{B_{max}}. \tag{1.15}$$

As recounted initially, the conservation of degrees of freedom entails that the amount of information required to exactly and completely specify the state at any given time remains constant: $\Delta C = \Delta S_{B_{max}} = 0$. The information required to exactly specify any global state vector consists of its coordinate values locating it in the overarching Hilbert space, $\mathcal{H}$.

> **Global Conservation of Degrees of Freedom (GCDF)**: The universe's total degrees of freedom are encoded in the basis vectors of its maximal Hilbert space, which has fixed dimensionality and contains the complete set of physically possible global state vectors.

GU is true if and only if the conjunction of GCDF and GUE is true: GU $\iff$ GCDF$\wedge$GUE. Now GCDF might hardly seem like an insight – more like a self-evident

truth. In general relativistic spacetimes, GCDF lays the foundation for global hyperbolicity because exhaustive foliation into Cauchy surfaces (i.e., complete spacelike hypersurfaces serving as 'global instants') conserves total degrees of freedom along a global time parameter. And GCDF itself is implied by GUE: GUE $\Longrightarrow$ GCDF. So a violation of GCDF certainly implies a violation of GUE: $\neg$GCDF $\Longrightarrow$ $\neg$GUE. That means the initial and final states must reside in overarching Hilbert spaces of different dimensionalities.

Consequently, a variation in degrees of freedom implies a violation of information conservation in terms of state specification. If new degrees of freedom appear, more information is required to exactly and completely specify the state. If some degrees of freedom disappear, less information is required to exactly and completely specify the state. A variation in degrees of freedom is a stronger violation of information conservation than deterministic dynamics or the effective constancy of coarse-graining because it's ontological, and information loss at this level implies weaker forms of information loss. When information is conserved, it's true that the direction of implication makes GUE stronger than GCDF. But when information is lost, the direction of implication makes $\neg$GCDF stronger than $\neg$GUE. Since the black hole information puzzle is about information loss, GCDF is a condition of unitarity to which we should pay attention.

Preservation of Entanglement: Last but not least, unitarity preserves the extent of entanglement with the environment. Hilbert spaces representing generic systems with pervasive internal entanglement tend to be minimally factorizable, which means that generic density matrices represent entangled subsystems (see Clifton and Halvorson 2001; Earman 2015). As I mentioned previously, the density matrix of a mixed state is underdetermined between macrostates and entangled subsystems. For that reason, Gibbs entropy has a twin in von Neumann entropy, S_{VN} – the entanglement-based incarnation of Shannon entropy (see Equation 1.16):

$$S_{VN} = -tr\rho \ln \rho = \sum_{i=1}^{n} p_i \ln p_i;\ p_i \geq 0,\ \Sigma_i p_i = 1. \tag{1.16}$$

Whereas Gibbs entropy is an uncertainty measure, von Neumann entropy is a correlation measure. Pure states have zero von Neumann entropy, which makes sense because they represent microstates of isolated systems, so they would not have information stored nonlocally in external correlations. Mixed states involving reduced density matrices have positive von Neumann entropy, which contain at least some information stored nonlocally in external correlations. Maximal entanglement comes about through a uniform probability distribution, thereby reducing von Neumann entropy to Boltzmann entropy.

Thermal states, which are of particular interest to us, are both thermodynamically and quantum mechanically coupled with their environment. Basic thermodynamic coupling involves equilibration and energy flow across subsystems to settle down to a uniform temperature. Quantum mechanical coupling obviously involves entanglement between subsystems. By stipulation then, aren't thermal systems entangled with their

environment? Quantum field theory answers in the affirmative, though a thermal system won't normally be entangled with all degrees of freedom in the external reservoir, given how vast that's supposed to be.

Partitioning a quantum field into a thermal subsystem coupled to a heat bath demarcates a causal boundary across spacelike separated regions, like at a moment in time across the event horizon. Unruh and Wald (2017) explain that "the entanglement between the field in two such causally complementary regions always occurs in quantum field theory, no matter what the spacetime or the (physically acceptable) state" (p. 2). Therefore, if a subsystem is entangled and thermal, its reduced density matrix traces out degrees of freedom of the heat bath, making it a thermal density matrix as well. As a result, thermal states unify Gibbs with von Neumann entropy, proving to be the key puzzle piece for black hole information loss.

Actually, what does entanglement have to do with information? As Esfeld (2004) remarks,

> The description in terms of improper mixtures therefore is an incomplete description of quantum systems in entangled states. It is not a description that refers to intrinsic properties of each quantum system (p. 604).

Whereas Gibbs entropy quantifies the information deficit contained within a system, von Neumann entropy S_{VN} quantifies the information deficit contained in an external system, I_{ext}, with a similar inverse relationship to negentropy and information (refer to Equation 1.17):

$$\Delta I_{ext} = -\Delta S_{VN}. \tag{1.17}$$

It should be no surprise given our analysis of Gibbs entropy that von Neumann entropy is invariant under unitary transformations: closed systems remain closed and entangled subsystems remain entangled – always to the same extent, might I add. Under non-unitary transformations, ΔS_{VN} and $-\Delta I_{ext}$ convey information loss as a transfer of nonlocal correlations to the environment.

Unitarity's preservation of entanglement illuminates the most basic assumption of all, without which none of the connotations of information conservation we've explored would even get off the ground. Information in our discussion has consistently boiled down to state specification and descriptive (in)completeness. Subsystems exhibiting external entanglement can only have descriptively incomplete, mixed states. Global systems lacking external entanglement by virtue of being global should always admit of descriptively complete, pure states.

> **Global Physical Statehood (GPS)**: All global states are informationally complete.

It's very intuitive to infer that ¬GPS necessarily follows from ¬GCDF, though that would be too quick. The number of degrees of freedom can in principle vary across global simultaneity slices but still represent the maximum at that instant if there's no external entanglement. Nevertheless, when GPS is violated nontrivially, i.e., when the

overarching Hilbert space cannot represent the maximum number of degrees of freedom and those remaining appear to be entangled with a fictional environment, then we're led to informationally incomplete, physically impossible global states, and we can be certain that GCDF was also violated. Due to the importance of entanglement in black hole evaporation, we also need to pay attention to information loss as $-\Delta I_{ext}$ and discern the status of GPS.

1.5 Candidates for Black Hole Information Loss

In this Section, I peruse the myriad of ways that black hole evaporation could violate the information conservation principles I've laid out. Even though it ultimately violates all of them in ascending order of severity, the most persuasive and impactful characterization of black hole information loss is the global appearance of external entanglement.

1.5.1 Red Herring 1: Thermodynamic Evolution

The most conservative candidate for non-unitarity in black hole evaporation is an effective violation at the level of coarse-graining, in which information loss appears to be of a familiar kind – the Second Law of Thermodynamics in action. There are a few proponents of this view, including Sonner and Vielma (2017); however, this strategy is a red herring because it calls for a radical change to Hawking's framework, and thus, an entirely different derivation of Hawking radiation.

Nonetheless, I wish to defend this strategy's conceptual plausibility in the absence of a rigorous justification in the literature (though it fails on technical grounds), and it's worthwhile to summarize how the Second Law could have potentially saved the day. The main punchline is the following: It's conceivable that we've been confusing an initial microstate with a nondegenerate initial macrostate, and therefore, conflating fine-grained pure-to-mixed evolution with coarse-grained pure-to-mixed evolution.

Hawking took very seriously the no-hair theorem, which states that black holes as seen from the outside are only distinguishable by a few parameters, namely mass, angular momentum, and charge, so the details of the collapsing matter are neglected. He also took very seriously the causal structure of the event horizon. If the collapsing matter is idealized as classical (where non-classical tunneling events are outright forbidden), the speed-of-light barrier prohibits any information about its microstate to escape after the formation of the event horizon. That information certainly can't come out as Hawking radiation because collapsing matter constitutes a localized system, whereas particle creation is a global process (Hawking, 1975). Though the black hole serves as a boundary condition in the evolution of the quantum fields, it never figures into the initial pure state.

Then why did Hawking and everyone else believe otherwise? The Fock vacuum is in fact a macrostate delimited by an expectation value of zero particle number and zero energy (so in a certain sense, zero temperature), yet it's also unique in Minkowski

spacetime, which is how Hawking treated past and future null infinity. As a usefully coarse-grained macrostate then, the Fock vacuum has zero Gibbs entropy. We assumed that it's only a microstate and pure-to-mixed evolution is nonsensical, but if we reconceptualize it as a macrostate, then we would realize that Hawking actually calculated a pure-to-mixed transition at the coarse-grained level. Violating the constancy of coarse-graining is expected when we go from one macrostate to another since unitarity doesn't apply to irreversible macrodynamics. Furthermore, Zurek (1982) shows that the von Neumann entropy of Hawking radiation (calculated from its density matrix) is significantly larger than the Bekenstein-Hawking entropy of the pre-evaporation black hole. This result presents strong evidence for the Generalized Second Law.

In order to have a clear-cut case of information loss as innocuous Gibbs entropy increase (but not von Neumann entropy increase), we'd need to prove that Hawking radiation is actually approximately thermal and perform purity-restoring corrections to Hawking's leading-order calculation. Unitarity at the fundamental level demands that a pure pre-evaporation state evolves deterministically into a pure post-evaporation state, which also insinuates that during late-evaporation stages, entanglement between the shrinking black hole and Hawking radiation decreases and eventually vanishes.

If we wish to demonstrate that what Hawking basically did was to add up the entropies of significantly smaller radiation subsystems that were, in fact, highly entangled with their immediate environment, then it would be sufficient to show that perturbative corrections to his calculations purify the thermal density matrix. Sonner and Vielma (2017) draw upon the Eigenstate Thermalization Hypothesis to demonstrate that pure states in internal thermodynamic equilibrium can very closely approximate exactly thermal states, and the corrections that would restore purity in the black hole evaporation case would be of order $e^{-S_{BH}}$, (where S_{BH} is the Bekenstein-Hawking entropy of the black hole).

However, Wallace (2020) points out that although the correlation factor will remain at $e^{-S_{BH}}$ for a long time due to the Poincare recurrence theorem, it will eventually become large again and deviate beyond perturbative corrections. Mathur (2009) refutes this strategy altogether by demonstrating that subadditivity constraints on entanglement between the black hole and radiation prevent sub-leading corrections from driving the radiation's entanglement down anywhere close to zero even when the black hole is of Planck mass. Put another way, he exposes the mathematical necessity for more substantive revisions to Hawking's calculation. We ultimately shouldn't be too surprised given that Hawking radiation started out as an entangled subsystem, so its thermality could very well signal that even post-evaporation, it's an improper mixture involving a subset of the totality of degrees of freedom.

1.5.2 Red Herring 2: Indeterministic Dynamics

The next available move is to reexamine how physical inferences are drawn from the formalism and broaden the mathematical objects that count as being descriptively complete, in alignment with GPS, to encompass density matrices. It's quite reasonable

then that the source of information loss lies in the dynamics. And if we're on the right track, black hole evaporation is pointing to entirely new, indeterministic laws of physics, which is exactly what Hawking (1976) proposed with his superscattering formalism. This interpretation of information loss as indeterminism has reverberated across academic circles and is the target of the mainstream narrative.

In this subsection, I aim to defy the reaction of the mainstream narrative: A metaphysical preference for determinism doesn't automatically make unexpectedly indeterministic evolution problematic, much less paradoxical. I also endeavor to expose the speciousness of casting information loss as indeterminism. So long as we think that the black hole information puzzle is about giving up GUE, we will justifiably but incorrectly assume that black hole evaporation is compatible with GCDF. In other words, there could exist an exact microstate corresponding to the post-evaporation Hawking radiation, but the indeterministic dynamics prohibit ever predicting it with certainty or even revealing which measurements would hypothetically allow an observer to experimentally discern the state. Yet such a picture lacks the force of a coherent physical interpretation in the face of lingering entanglement and the one-way destinations that are black holes.

First and foremost, what needs to be spelled out is why pure-to-mixed evolution is indeterministic. Hawking (1976) demonstrated that the value of the initial state does not imply the value of the final state in black hole evaporation. No matter what the initial state vector, we end up with a thermal density matrix and straightforwardly lose information about amplitudes along the evolution. Hawking called this connotation of information loss, which follows from the absence of a nondegenerate observable, the breakdown of predictability. Furthermore, since the thermal density matrix no longer contains information about the amplitudes to recover the initial state vector, the time-reversed evolution suffers from a failure of retrodictability. This is what he seems to have in mind when he says,

> This result can be regarded as a quantum version of the "no hair" theorems because it implies that an observer at infinity cannot predict the internal state of the black hole apart from its mass, angular momentum, and charge: If the black hole emitted some configuration of particles with greater probability than others, the observer would have some a priori information about the internal state(Hawking, 1976, p. 2462).

Susskind (2008), in his popular book, *The Black Hole War: My Battle with Stephen Hawking to Make the World Safe for Quantum Mechanics*, laments the dynamical irreversibility of black hole formation and evaporation.

> [Hawking's] view was that the precise details of the gas cloud–whether it was made of hydrogen, helium, or laughing gas–would go down the drain, past the point of no return, and disappear with the black hole when it evaporated...Reversing all the final particles and letting the whole thing run backward would not reconstruct the original input. According to

> Hawking, the result of reversing the final radiation would just be more undifferentiated Hawking radiation (Susskind, 2008).

Susskind and mostly everyone who've followed Hawking's lead thought that the source of information loss lies in the manifestly indeterministic dynamics. Details are washed out when running the laws forward in time, and time-reversal invariance is a bygone relic of theories past. Nevertheless, if that's all there is to the story, then there's no prima facie reason for information loss to be puzzling. In fact, we should resist framing the black hole information puzzle in a way that prematurely adjudicates on the measurement problem.

Hawking (1976) gives the impression of being committed to GCDF when he states that information about the black hole interior is ontologically present in the Hawking radiation despite the black hole's disappearance, intimating that degrees of freedom are conserved from the pre-evaporation to post-evaporation state. He clarifies,

> Of course, if the observer measures the wave functions of all the particles that are emitted in a particular case he can then a posteriori determine the internal state of the black hole but it will have disappeared by that time (Hawking, 1976, p. 2462)

With a more poetic undertone, Hawking (1976) muses,

> Einstein was very unhappy about the unpredictability of quantum mechanics because he felt that "God does not play dice." However, the results given here indicate that "God not only plays dice, He sometimes throws the dice where they cannot be seen" (p. 2464).

God throws the dice, they land on determinate faces with some probability, but they're thrown where they cannot be seen. Notice the tension here. There could be an exact microstate, perhaps a pure state of Hawking radiation, that would indeed expose the past internal state of the black hole. Page (1995) understood Hawking to be saying just this; he was told that the final density matrix is an "intermediate tool" and not "literally the actual final state of the system" (p. 7). Rather, the final density matrix should be used to calculate conditional probabilities of measuring a final pure state given an initial pure state. Page (1995) continues, "Then the asymmetry may indeed be more in the conditional nature of the probability than in any time asymmetry (e.g., CPT noninvariance)" (p. 7).

The mention about CPT (i.e., charge, parity, time) invariance is crucial. If black hole evaporation does not truly violate CPT invariance, then the dimensionality of the final overarching Hilbert equals that of the initial overarching Hilbert space, and degrees of freedom are conserved in the superscattering formalism just as I had intuited from Hawking's quotes. Therefore, the only levels at which information is not conserved are the first two – thermodynamic evolution and indeterministic dynamics. In fact, it may not even be the case that unitary evolution (GUE) is violated. Page

(1980) has argued that if black hole formation and evaporation is compatible with a CPT-invariant superscattering operator, then it's also compatible with a regular, unitary scattering operator mapping pure initial states to pure final states.

The problem with conserving maximal Boltzman entropy/information storage capacity is that GCDF undermines the event horizon as a strict causal barrier. The entangled negative-energy quanta in the black hole would have had to surpass the speed-of-light-barrier on a much larger scale than low-probability tunneling events along classically forbidden trajectories. Yet despite biting that bullet, it's strange that nature would conspiratorially prevent that information from being accessible via the laws. This position leaves us with a bizarre contradiction that the event horizon concurrently is and isn't a causal barrier. Information can't escape, but degrees of freedom can.

What's also odd is that Hawking (1976) interpreted pure-to-mixed evolution as a form of metaphysical indeterminacy over and above the uncertainty principle, but only with epistemological consequences completely disassociated from the ontology. I am at a loss to understand what's different about interpreting the final thermal density matrix as a novel type of global state as opposed to a statistical ensemble with missing information about the amplitudes of the unknown yet ontologically meaningful final pure state, in which case we'd encounter the very objections we ran into in Section 1.5.1. Furthermore, the form of the new indeterministic laws remains opaque. Hawking's superscattering formalism might be an immediate contender, but it isn't of much help because it relates ingoing density matrices to outgoing density matrices, whereas the intermediate evolution is shrouded in a black box (or should I say, black hole!).

Let's see if anyone else can do better. If a more comprehensive set of dynamics can explain what's going on during black hole evaporation, then save for an a priori metaphysical preference, there's no reason to render information loss undesirable. The funny story is that quantum mechanics already has non-unitary variants – collapse theories. Okon and Sudarsky (2017) argue that the condition of unitarity in the black hole information loss paradox and the condition of unitarity as one of three horns of the measurement problem trilemma are in fact the same condition; therefore, denying it solves two problems at once.

Inspired by Penrose's proposal, Okon and Sudarsky (2014) believe that stochastic collapse – the rate of which accelerates in the high-curvature regimes of black hole interiors and near-singularity regions – satisfactorily accounts for information loss. Given this picture, I'm going to venture out and assert that it's no wonder Hawking calculated a thermal density matrix for the radiation state. The stochastic collapses along the way would have prohibited the prediction of a unique final state, so the most we could hope to learn are probabilities for the most-likely projective outcomes. But while one sense of information may be lost due to indeterministic dynamics, another sense is not lost due to the time-averaged constancy of coarse-graining.

> [Penrose] argues that the information lost into the black hole causes trajec-

> tories in phase space to converge and volumes to shrink. That is because different inputs give rise to the same output. He holds, however, that this loss of phase space volume is balanced by the quantum spontaneous collapse process since, in the quantum case, several outputs may follow from the same input. (Okon and Sudarsky, 2014, p. 140).

Taking this result at face value, indeterministic dynamics produce variability in coarse-graining over the short-term. Liouville's theorem, in which the phase space volume of the initial macrostate is conserved at every infinitesimal time interval, is certainly not upheld. Yet over sufficiently long time periods, average phase space volume is conserved, which is why Okon and Sudarsky (2014) argue that supplementing Hawking's framework with stochastic collapse satisfactorily restores information for all practical purposes. What they have not addressed is how we're justified in representing the pre- and post-evaporation states in the same phase space with fixed dimensionality. A large part of the mystery of information loss has to do with the post-evaporation status of black hole degrees of freedom since the thermality of late-time Hawking radiation is indicative of an improper mixture.

This mystery is just as pronounced in a Hilbert space construction. After identifying a Hilbert subspace $H(t_0)$ for an initial macrostate, whose orthonormal basis vectors are eigenvectors of relevant macro-observables, and thus, constitute a proper mixture of possible pure states, non-unitary evolution will vary the dimensionality of that subspace after each stochastic collapse, thereby affecting the constancy of coarse-graining in the short term. Stochastic collapse is a discontinuous projection of a generic unit vector onto one of the orthonormal basis vectors of the overarching Hilbert space, $\mathcal{H}$. The amplitudes of the spectral decomposition of the initial pure state probabilistically predict projective outcomes according to Born's rule. Due to indeterministic dynamics, multiple initial pure states can be projected onto the same basis vector, and on the flip side, the same initial pure state can also be projected onto multiple basis vectors. Therefore, similarly to the phase space representation, the time-averaged dimensionality of $H(t)$ remains constant since the many-to-one and one-to-many dynamical histories tend to balance out.

But Okon and Sudarsky's account doesn't work unless the dimensionality of the overarching Hilbert space, $\mathcal{H}$, is fixed, and to see why, we benefit from the Bloch hypersphere visualization. Any initial pure state is a unit vector intersecting the Bloch surface. However, the vast majority of stochastic collapses will project it onto a different state and reduce its length, transforming it into an interior Bloch vector representing a mixed state. For this reason, retrodictability fails since the shortening of the original unit vector represents a loss of information about the initial amplitudes. Renormalization into a unit-length pure state cannot recover that information, so the washing away of initial conditions is permanent.

Nonetheless, what's imperative to keep in mind is that interior Bloch vectors are always proper mixtures, and regardless of how the ensemble of $H(t)$'s basis vectors changes over time, it's fair game for stochastic collapses to project onto any one of $\mathcal{H}$'s

basis vectors for Okon and Sudarsky's argument about the net constancy of coarse-graining to go through. While projection screens off degrees of freedom, it doesn't outright remove degrees of freedom as does a partial trace. Consequently, $\mathcal{H}$'s degrees of freedom have to be conserved, presupposing that all global states, like the post-evaporation radiation state, can access the full set of degrees of freedom with the black hole included.

Therefore, Okon and Sudarsky haven't convinced me that indeterministic dynamics explain and ameliorate information loss due to the same objection I levied against Hawking. The main issue with exactly thermal Hawking radiation is that we can't represent the final state as a microstate in its own right without major modifications to the kinematics, even if we attempt to expand the representation of descriptively exhaustive states as mixed density matrices. The cleanest interpretation of mixed density matrices in this context is that of macrostates involving ensembles of collapse histories.

Though stochastic collapses certainly account for why we can write only the initial state as pure and all succeeding states as density matrices, that limitation stems purely from the lack of predictability and retrodictability. The indeterministic dynamics do not preclude the description of a radiation microstate with more fine-grained information since any collapsed state that's less than unit length is renormalizable into a pure state. The epistemically inaccessible yet ontologically meaningful dynamical history of the actual universe exclusively involves microstate-to-microstate (i.e., pure-to-pure) evolution, albeit discontinuously in the relevant state space. Therefore, the arguments of Okon and Sudarsky about dynamical collapse proposals solving the puzzle are question-begging. The question revolves precisely around the physical interpretation of mixed post-evaporation states, which they assume from the outset.

Nevertheless, what this analysis has indeed inspired is that the nature of the dynamics is somewhat irrelevant to the puzzle precisely because determinism and indeterminism are both compatible with the global conservation of degrees of freedom (GCDF). Global unitary evolution (GUE) implies GCDF but not the other way around. That's why it was prudent to decouple the two metaphysical claims embedded in unitarity about the nature of laws and the quantification of fundamental ontology as separate conditions for information conservation. From that perspective, unitarity as a package deal may be dispensable, but one of the necessary conditions for unitarity, GCDF, lies at the heart of the puzzle.

1.5.3 Golden Egg 1: Elimination of Degrees of Freedom

To recap, the indeterminism baked into pure-to-mixed evolution looks nothing like the stochastic collapse of the wavefunction, where we go from one pure state to another through a somewhat ad hoc transition. The conclusion of the mainstream narrative is too quick and refuses to follow to its logical end the implications of black holes as regions of no escape until the very end of their lives. The mainstream narrative obscures the realization that pure-to-mixed evolution is indeterministic because of

ontological instabilities in the 'quantity of being'. Not all flavors of indeterminism are created equal.

The flavor of indeterminism that overpowers all the rest is the ad nihilum destruction of degrees of freedom, which occurs if we take the singularity seriously as an edge of spacetime that ends worldlines. It's clear that Susskind, one of the most prominent voices in the debate, was worried about this but refused to express it in scholarly articles, so his official line was to narrowly construe unitarity as invertibility. However, in his book targeted towards general audiences, where he wanted to sway the masses into taking his side over Hawking's, he emphasizes the implications of the singularity on collapsing matter and negative-energy Hawking particles.

> When a black hole evaporates, the trapped bits of information disappear from our universe...It is irreversibly, and eternally obliterated (Susskind, 2008).

A major consequence of the elimination of degrees of freedom is that black hole evaporation cannot be modeled in a fixed-dimensional state space, whether that's Hilbert space or a more exotic upgrade. The dimensionality of the post-evaporation Hilbert space is smaller than the dimensionality of the pre-evaporation Hilbert space. Therefore, the logarithm and overall entropy of the entire state space has decreased. As I alluded to in Section 1.5.1, it's no wonder that final Hawking radiation is thermal and in an improper mixture. It's still only a proper subset of what should've been the full, global set of degrees of freedom.

Identifying how information is lost in this scenario becomes a subtle exercise. According to the expression for negentropy in Equation 1.12, information has actually been gained, not lost, because a decrease in the ensemble size corresponds to a decrease in uncertainty. However, I want to resist the application of negentropy at this juncture. The utility of information-theoretic entropy is to quantify the degrees of freedom that remain independent under epistemically and physically salient gross constraints. Information gain or loss over time provides valuable insight into how the gross constraints are changing, as with the Second Law of Thermodynamics. However, no gross constraints are considered when enumerating a system's maximum independent degrees of freedom, so changes in that quantity provide vacuous insight from an information-theoretic standpoint.

A much more pedestrian and common sense conclusion is that we've lost all information associated with the discarded degrees of freedom. We need to play a modified game and employ Susskind and Lindesay's conceptualization of maximal Boltzmann entropy as information storage capacity. According to Equation 1.15, the final radiation system has less information storage capacity than the initial vacuum coupled to the black hole metric, leading to $-\Delta I_{max}$. Unlike the inverse relationship between Gibbs entropy and available information, the decrease of maximal Boltzmann entropy really does correspond to information loss because the discarded degrees of freedom will never be accessible, even in principle, either epistemically or physically. The most

information we could hypothetically retrieve is the reduced storage capacity reflected in the smaller Hilbert subspace.

The opposing deductions from equations 1.12 and 1.15 arise when we're taking the entire universe as the system of interest. It's unclear how to make dynamical sense of changes in negentropy and storage capacity without an environment to inject or absorb degrees of freedom. It's even debatable whether physics as an enterprise could continue to be carried out because it's hard to surmise what would even count as a law of nature. After all, degrees of freedom are free variables, the values of which serve as input into the dynamical laws. It's questionable whether laws could coherently link states specified by different variables. In a striking moment of passion, Susskind (2008) captures the burden of the black hole information puzzle:

> All hell would break loose in all of physics, not just black hole physics, once the door to information loss was opened (Susskind, 2008).

The history of modern physics has taken for granted that whatever the fundamental ontology, conservation laws point to profound metaphysical truths about the underlying stability of the material world. Whether it's energy conservation or energy-momentum conservation or charge conservation or another invariant across time as given by a particular theory, there's a deeply ingrained assumption that 'quantity of being' is neither created nor destroyed.

I'd like to now preempt potential skepticism that black hole evaporation is radical in its implications for the ebb and flow of quantity of being. One may push back that predecessor theories have already set a precedent for variation in the fundamental ontology. After all, Fock space accommodates indeterminate particle number. The conservation of degrees of freedom as the unchanging dimensionality of a mathematical state space is an artifact. It's too easy to make the presence or absence of a supposedly material entity an abstract axis in a state space, so why not just do away with the intermediate step?[7]

And maybe an appeal to what acceptable laws are is not convincing either. We once thought that only Euclidean space was compatible with there being laws of nature, or a universal time parameter. In reasoning that the non-conservation of degrees of freedom is inhospitable to laws with which we're already familiar, i.e., maps between states, it might sound like a metaphysical bias is bleeding into proclamations about what's physically possible (see e.g., Adlam 2022). Maudlin et al. (2020), for instance, bemoan that Aristotelian notions about the eternity of substances have already bled into the connotation of conservation laws.

To adjudicate on the acceptability of black hole information loss as the elimination of degrees of freedom based on metaphysical considerations depends on one's stance about the role of metaphysical theorizing in physics. If one subscribes to hardcore naturalized metaphysics (see Ladyman and Ross 2007), then information loss as the

[7]I'd like to thank Tim Riedel for discussion about representation relations between mathematical degrees of freedom and material entities.

elimination of degrees of freedom could be a metaphysical revolution, not a problem. Physics is ultimately a guide to metaphysics, and if we have to strip ourselves of our pre-scientific beliefs, then so be it.

I don't have knock-down rebuttals to such objections, especially without getting into the weeds of a particle versus field ontology (see e.g., Fraser 2022) or stepping into the landmine of candidates in metaphysics for the fundamental ontology (see e.g., Tahko 2018). Nor do I wish to disenfranchise dynamical collapse approaches that patently infringe upon conservation laws, such as for energy. But I do want to caution against too liberal a reading of the formalism.

Granting that even a fixed-dimensional state space admits of indeterminacy or variation in the quantity of being, I nonetheless emphasized in the previous section that all degrees of freedom are accessible throughout the evolution. The adage of physical stuff neither being created nor destroyed graduates to the more mature qualifier of 'at least not permanently'. So, particles may be created and destroyed in a Fock space construction with non-unitary dynamics, but they can always be created again and they can always be destroyed again. Therefore, conservation laws encode statistical averages or expectation values.

Black hole evaporation, however, is unprecedented because it seems to drop the qualifier of 'at least not permanently' by invoking a permanent partial trace. As I stated earlier, the reduction in maximal Boltzmann entropy simply exiles degrees of freedom from the physical realm. Given Page's colorful reaction about this "rather violent violation of CPT invariance" (Page, 1995, p. 5), nonchalance towards Hilbert space variability is an untenable attitude without checking empirical adequacy.

Some physicists, including Banks et al. (1984), worry that non-unitary evolution of the superscattering variety admits of egregious time-dependent violations to energy conservation even with asymptotic boundary conditions. They further argue that it's inconsistent with empirical evidence favoring unitary (or near-unitary) processes in particle accelerators. For example, calculations for scattering experiments in quantum field theory involve a multiplicity of possible intermediary virtual processes, one of which could be the formation and evaporation of microscopic black holes (see e.g., Wallace 2020). Other physicists, however, including Unruh and Wald (2017), counter that energy conservation is not violated if the black hole "retain[s] a 'memory' of what energy it previously emitted" (p. 13). Moreover, deviations from unitarity in low-energy laboratory experiments are anticipated to be unobservable and therefore, negligible. So for now, we've arrived at an impasse.

In the face of disagreements about whether violations to conservation laws can ever be settled experimentally, as well as the lack of empirical confirmation for Hawking radiation, naturalized metaphysics is certainly not beholden to the loss of global information storage capacity. Rather, it's silent on the question, and Jaksland (2023) argues that naturalized metaphysics is actually unfit to handle a situation without "sufficient epistemic warrant" (p. 2). That of course doesn't license putting theoretical physics on hold. Alternatively, if one subscribes to mutual interdependence between physics and metaphysics, a more reasonable stance in my eyes, when sometimes metaphysics serves

as a lighthouse for physics (see French and McKenzie 2012), then a meta-inductive argument should be sufficiently persuasive that eliminating degrees of freedom is a radical proposition.

Notwithstanding, we haven't touched upon the role of entanglement in black hole evaporation, and it turns out that the decrease of maximal Boltzmann entropy is not the sole form of information loss in this puzzle. Thermal systems are normally entangled systems, and if the final radiation state was insensitive to the disappearance of degrees of freedom inside the black hole, then it might be insensitive to updating its correlation structure.

1.5.4 Golden Egg 2: Appearance of External Entanglement

Thus far, we've established that black hole evaporation involves pure-to-mixed evolution because Hawking radiation excludes degrees of freedom that got trapped inside the black hole. We've examined this story through the lens of thermodynamics, indeterministic dynamics, ontological instabilities, and finally, we're going to do so through entanglement. Pure-to-mixed evolution is accompanied by an increase in von Neumann entropy, so the only remaining explanation is to concede that Hawking radiation is entangled with another subsystem, which has ramifications for the amount of information stored in external correlations and foreshadows an impending metaphysical upheaval.

Semi-classical gravity intimates that the most sensible, *immediate* ontological interpretation of thermal Hawking radiation is that of a global entangled system. It's a global system because of the violation of GCDF – a proper subset of degrees of freedom have been obliterated by the black hole singularity. But because thermal systems have positive von Neumann entropy equivalent to their Gibbs entropy, the final Hawking radiation also seems to be an entangled system.

What if the extant degrees of freedom aren't exactly thermal? If we were to entertain a decrease in maximal Boltzmann entropy, then perturbative corrections to Hawking's framework could plausibly purify Hawking radiation as per the Eigenstate Thermalization Hypothesis without having to restore black hole degrees of freedom. However, Hawking's intuitions would then be dashed that measuring the wavefunction of post-evaporation radiation would allow for an a posteriori determination of the black hole interior despite its disappearance by that time. The only medium for gaining information about the black hole after the fact is through entanglement.

We know that during evaporation, Hawking radiation had been entangled with modes inside the black hole. Therefore, prior knowledge of the overall pure state, namely the nonlocal trans-horizon correlation structure, would be conducive to gleaning the exact internal state of the black hole simply by measuring the exterior radiation state. The same procedure presumably applies to post-evaporation measurements, based on what Hawking (1976) has in mind. Final Hawking radiation must somehow still be correlated with past interior degrees of freedom to reveal the values of bygone black hole observables. The persistence of entanglement with annihilated degrees of

freedom thus raises the notion of 'nonlocal' to a whole new level.

On the face of it, this situation shouldn't be too surprising given that entanglement is very difficult to get rid of in quantum field theory. Clifton and Halvorson (2001) prove that no local operations, unitary or non-unitary, can remove entanglement in quantum field theory. Therefore, an appeal to curvature-induced collapse à la Okon and Sudarsky (2014) would be futile to sever the entanglement. Even if the high-curvature regime near the black hole singularity were to induce collapse for the quantum states of interior degrees of freedom, such as the negative-energy partners of Hawking pairs, the positive-energy partners would remain untouched and incur no modifications to their state, thereby retaining positive von Neumann entropy.

All things considered, global states should not have any information stored in external correlations because there's nothing outside to interact with, and the purity of the initial state reflects that. However, the impurity of the final state indicates the spontaneous appearance of entanglement with an unknown environment. According to Equation 1.17, the increase in global von Neumann entropy signals information loss. Information that used to be self-contained is now stored in unknown external correlations.

As far as I know, Page (1995) is one of the few interlocutors who had overtly characterized information loss as the increase in von Neumann entropy of the whole system, pinning it down as the relevant connotation for black hole evaporation. However, he didn't exactly elaborate on red flags other than doubting the utility of semi-classical gravity as an approximation. His takeaway was that information loss comes about from treating mascroscopic black holes classically when instead they should be treated quantum mechanically from the outset.

Unfortunately, Page's insight has been drowned out by the mainstream narrative conflating information loss with indeterminism. As I mentioned in passing in Section 1.4, globally hyperbolic spacetimes are compatible with deterministic dynamics, such as unitary evolution, because they conserve degrees of freedom. They do so by virtue of being exhaustively foliable by Cauchy surfaces, i.e., all-encompassing simultaneity slices. For that reason, Belot et al. (1999), Maudlin (2017), and Wallace (2020) argue that we just need to pose a single question to elucidate the source of information loss. Are evaporation spacetimes globally hyperbolic? Hawking's answer is no. They then proceed to deem the mystery about pure-to-mixed evolution solved and dismiss any further concerns about the evaporation-time puzzle being paradoxical. For philosophers at least, myself included, indeterminism in itself is not a big deal, which would explain their nonchalance. Maudlin (2017) goes so far as to avow that over half of his paper is redundant upon recognizing the role of global spacetime structure.

> That means that there never has been any grounds to expect the transition to be either retrodictable or unitary. Quantum theory does not imply, and never has, that such a transition must be retrodictable or unitary or "preserve information". There were no grounds in 1975, and remain no grounds today, to expect information to be preserved. There was never

> any information loss paradox based in fundamental properties of quantum theory and relativity. This paper could end here (Maudlin, 2017, p. 10)

But Maudlin's paper didn't end there. Professing as much flippantly downplays his own realization that there's more to information loss than non-global hyperbolicity, especially when his next paragraph literally reveals the need for additional questions to frame the puzzle.

> What we have left to us, then, are two important questions. First, if the information about what originally formed the black hole (and more generally whatever passed through the event horizon) is not present on [the post-evaporation state], where is it? And second, how did so many prominent and brilliant physicists manage to get so confused? The answers to these questions may be linked.

Now there's the jackpot question...where is the information on the post-evaporation state? Maudlin doesn't clarify why we should even be asking this question, and as I've endeavored to show, the mainstream narrative about indeterminism doesn't steer us in this direction. So indeed, many "prominent and brilliant physicists" did "manage to get so confused", but not because, as Maudlin (2017) declares, "*forty years of effort has been directed at a non-problem*" (p. 10). If that were true, then Maudlin would also be "directing effort at a non-problem", but he's too incisive to fall for that trap when he's actually uncovered a real problem. Most scholars have confusedly been responding to the wrong problem. They've either been dismissing the non-problem or attempting to tackle the real problem while mistakenly believing it's the non-problem.

It's predominantly Page (1995) who has strewn a trail of breadcrumbs towards the real problem. He noticed that the initial and final Hilbert spaces have different dimensionalities, so the information of discarded degrees of freedom, i.e., the gap in Boltzmann entropy, must've gone somewhere. He also made us aware that global von Neumann entropy has increased during evaporation, so late-time Hawking radiation must be entangled with a complementary subsystem that again, must be somewhere. Maudlin (2017) doesn't overtly make the connection between the location of missing information and unaccounted for entanglement, but reading between the lines, that was the reason behind his paper not ending where he said it could.

Consequently, the conundrum of the spontaneous appearance of external entanglement invites the question: Is final Hawking radiation a global state or not? This right here is the central puzzle of black hole information loss.

> **Central Puzzle (CP)**: With which system is late-time Hawking radiation entangled?

With no blatant prescription from semi-classical gravity about how to locate the complementary entangled subsystem, GPS is violated since post-evaporation radiation states are informationally incomplete. The forceful bite of the black hole information

loss puzzle is that evaporation leads to what's probably a physically impossible global state – one registering entanglement with unphysical degrees of freedom. Let's bring the problem into sharper relief by using a Hilbert-space formulation for visualization.

The pure state of a black hole and its surroundings starts out as a vector in a well-defined, overarching Hilbert space. During evaporation, entanglement between the black hole and Hawking radiation grows until the final evaporation event, meaning that neither of their associated subspaces can be factored out from the overarching Hilbert space. Then after evaporation, the black hole interior degrees of freedom are eliminated from the overarching Hilbert space and its dimensionality drops. But the radiation subspace's positive von Neumann entropy implies that it still feels entanglement with the black hole subspace. Because the black hole subspace couldn't have been factored out with the persistence of entanglement, the concurrent violations of GCDF and GPS imply that it was illegally excised, rendering the final Hilbert space pathological.

My chief suspicion underlying the black hole information puzzle is whether entangled Hawking radiation could even be a legitimate global state. In order to solve this problem, we could embed the pathological Hilbert space into a larger, well-defined Hilbert space with additional degrees of freedom. This move is equivalent to manually 'stitching' another subspace back into the pathological Hilbert space to purify the radiation subspace and recover the initial dimensionality/Boltzmann entropy, the most obvious candidate being the black hole subspace. The storage capacity of this curative subspace must match the von Neumann entropy of the radiation subspace, thus forging an unexpected equality between equations 1.15 and 1.17.

However, as I mentioned in Section 1.5.1, the von Neumann entropy of Hawking radiation is greater than the Bekenstein-Hawking entropy of the initial black hole. Without a better foundation from quantum gravity about the appropriate interpretation of Bekenstein-Hawking entropy, the microscopic structure of black holes, and the dynamical mechanism behind Hawking radiation, we may not be able to get away with simply stitching the black hole subspace back in. In fact, because the final, reduced Hilbert space was supposed to be the new overarching Hilbert space, albeit with diminished information storage capacity, this stitching move is tantamount to introducing unphysical degrees of freedom to compensate for the missing ontology.

1.6 Conclusion: Is Black Hole Information Loss Paradoxical?

We've arrived at the million-dollar question: Is black hole information loss truly a paradox or just an unpalatable consequence of applying QFT to the one-way destinations that are black holes? I've endeavored to show that the unproblematic versions of information loss, the Second Law of Thermodynamics and indeterministic dynamics, fall short in explaining black hole evaporation, whereas the problematic versions of information loss, a reduction in degrees of freedom and the appearance of external entanglement for a global system, are adjacent to being physically and metaphysi-

cally unintelligible. The cherry on top is that my focus on entropy, particularly von Neumann entropy, captures the stronger forms of information loss using the most appropriate information-theoretic tools in the field.

Now accepting these stronger forms of information loss is prima facie too high a cost without exploring other options, especially taking into account that semi-classical gravity is not a final theory. Nonetheless, as per Sainsbury (2009), a paradox involves a problematic conclusion derived from plausible premises through legitimate reasoning. In order for black hole information loss to constitute a paradox, it must definitively be established that information loss is indeed a problematic conclusion involving an explicit contradiction. In Chapter 2, I will demonstrate that an entangled global system cloaks a contradiction I term 'phantom entanglement'.

Chapter 2

The Phantom of the Space Opera: Why Black Hole Information Loss is *Really* Paradoxical

2.1 Introduction: The Grown-Up Answer

In Chapter 1, I employed information-theoretic resources to argue that the black hole information loss puzzle, as originally conceived by Hawking (1976) and dubbed by Wallace (2020) the "evaporation-time paradox", raises red flags and deserves further scrutiny regarding its potentially paradoxical status. Despite the fact that it has been one of the most thought-provoking, furor-inducing, and trendiest problems to work on in contemporary theoretical physics, a recurring theme in philosophy of physics has been to dismiss it as unworthy of pursuit. The angst displayed by many physicists over information loss, several philosophers aver, is the result of sociological forces distorting a perfectly reasonable conclusion that black holes are supposed to be one-way destinations.

> [T]he controversy can be cast as a clash of sub-cultures in physics, with the high energy physicists typically eager (if not desperate) to avoid the paradox, while the general relativists are generally more prepared to embrace it (Belot et al., 1999, p. 191).

> There is no "information loss" paradox. There never has been. If that seems like a provocation, it's because it is one. Few problems have gotten as much attention in theoretical foundations of physics over the last 40 years as the so-called information loss paradox... Probably no completely satisfactory non-sociological explanation is possible (Maudlin, 2017, p. 2).

The main objection has been that Hawking uncovered not a paradox so much as a knowledge deficit to be filled in by a future theory of quantum gravity. For quite some time, the official line from physicists has been to postpone even the framing of

the paradox. Susskind and Thorlacius, among the foremost champions of information conservation, hedged their bets back in 1994.

> We conclude that the information paradox can only be precisely formulated in the context of a complete theory of quantum gravity and that the issue of information loss cannot be definitively settled without such a theory (Susskind and Thorlacius, 1994, p. 966).

Numerous physicists and philosophers have taken advantage of this epistemic purgatory to embrace the possibility of post-evaporation information loss without proffering a positive account for what such a physical picture entails. Or if they've chosen to remain agnostic, they've taken shelter in what I label the "grown-up answer": It's futile to critically assess the status of the black hole information loss puzzle or even bother spelling out what the paradox is supposed to be without a final theory of quantum gravity.

The grown-up answer only makes sense under the pretense that Hawking's semi-classical framework of black hole evaporation is an unsuitable target of analysis. Advocates of the grown-up answer tend to be steeped in quantum gravity research anyway, so why dabble in a framework susceptible to countless weaknesses? After all, semi-classical gravity is not so much a full-fledged theory as it is a first pass at coming up with a unified theory of quantum gravity. It smashes together quantum matter fields and classical curved spacetimes, forcing the parent theories to play nice, at least temporarily. More tellingly, Hawking's derivation of an evaporation spacetime is neither an exact solution nor an approximation of a solution nor a family of solutions. It's what Curiel (2020) calls a "principle model": the result of complex arguments based on subtle, often imprecise and unproven assumptions.

It's these so-called weaknesses that make Hawking's semi-classical framework conducive to fruitful analysis. Isolating subtle, imprecise and unproven assumptions that we did not realize were contradictory and may have carried over into quantum gravity approaches is exactly what we need to do to make progress. In the best-case scenario, if a critical assessment of the black hole information loss puzzle yields a genuine paradox, then we have an actionable prescription for quantum gravity as well as a definitive benchmark to evaluate the plethora of live proposals advertising that they've solved or are on their way to solving it. My topmost aim in this chapter is to deliver on the best-case scenario. As Belot et al. (1999) remark:

> Whether information loss is an unexceptional consequence of global spacetime structure or a harbinger of the disintegration consequence of tractable microphysics is an important question meriting detailed examination (p. 203).

In that vein, I accept the invitation to examine and defend how the black hole information loss puzzle is indeed a "harbinger of the disintegration of tractable microphysics" that's still relevant to the current debate. No proposal responding to the mainstream narrative of black hole information loss that's worth its salt can sidestep

dealing with the contradiction I'm about to expose. The predominant ones in the discourse are already situated in various quantum gravity approaches and have unintentionally dealt with this contradiction without ever having acknowledged it. They thought they were restoring unitarity when in fact, they were confronting what I've branded the "paradox of phantom entanglement".

Although the definition of a paradox isn't set in stone, my chosen strategy is to show that post-evaporation information loss is a *prima facie* unacceptable conclusion following from *prima facie* acceptable premises (see e.g., Sainsbury 2009). These prima facie acceptable premises comprise two separate arguments having to do with the dynamical narrative of black hole evaporation, one from general relativity and the other from quantum field theory. Each argument results in a prima facie acceptable conclusion individually, but by combining the conclusions, I demonstrate a contradiction bearing on a kinematic aspect of black hole evaporation, which I call the paradox of phantom entanglement. Because the paradox arises from competing predictions of general relativity and quantum field theory, it's a gripping catalyst to make strides in quantum gravity research.

This chapter is organized as follows. I update and add to the methodological commitments of Chapter 1 in Section 2.2. In Section 2.3, I summarize the red flags of the black hole information loss puzzle and bring them into sharper relief by analyzing evaporation spacetime structure. In Section 2.4, I lay out two arguments leading to my formulation of a powerful paradox in Section 2.5: 1) the general relativistic argument, which establishes that late-time Hawking radiation constitutes a global system; 2) the quantum theoretic argument, which entails that late-time Hawking radiation makes up an entangled subsystem; and 3) the paradox of phantom entanglement, in which late-time Hawking radiation is entangled with unphysical degrees of freedom. A resolution to the paradox of phantom entanglement mandates modifications to Hawking's original framework. In Section 2.6, I explore big-picture strategies towards a resolution, for which input from quantum gravity is indispensable.

2.2 Methodological Commitments

My focus in this chapter is to dispel confusion over the underlying physical picture of black hole information loss. To jumpstart physically intuitive explanations relevant to the paradox of phantom entanglement, it's worthwhile to build upon prior methodological commitments, specifically the legitimacy of the states-plus-laws toolkit as well as the interplay between entanglement and information conservation. There's no getting around the technical jargon in such a high-level, complex debate, but to keep the analysis accessible and concise, I intentionally provide mostly qualitative expositions of key vocabulary, sidestepping their mathematically rigorous definitions.

2.2.1 States-Plus-Laws Toolkit

Upgrading the black hole information loss puzzle to a paradox hints at an impending pressure point in our semi-classical states-plus-laws toolkit. Something will have to give, whether it's the categorization of admissible states, the choice of dynamical laws, the spacetime structure within which we're attempting to utilize the toolkit, or the toolkit altogether. Throughout the remainder of the discussion, we will be analyzing these pressure points, so as a prelude, let's review some basic physical implications.

The invocation of global states presupposes that spacetime can be stratified into global simultaneity slices, i.e., universal moments, a non-trivial feat in general relativity. Only spacetimes that are time-orientable and foliable everywhere, where local future and past lightcones are more or less aligned, admit of global states (see e.g., Manchak 2011). Granting this presupposition for evaporation spacetimes, global states thus represent universal snapshots of the ontology.

But we're interested in compiling snapshots into feature films to ascertain the conservation or loss of state properties during black hole evaporation. Therefore, we must be able to compare states. To put it more bluntly, we require laws whose job it is to provide predictive and/or retrodictive algorithms relating one state to another. As soon as we relate two or more states, we're in the business of chronicling the target system's dynamical evolution.

Finally, symmetries of dynamical laws offer shortcuts to ascertain conserved quantities, including various information measures. We also need the right kind of symmetries to ensure that different foliations don't give rise to physically distinct spacetimes, otherwise, we'd be tampering with relativity. While snapshots across foliations capture different simultaneity slices of the ontology and are ordered in different successions, the feature film of a relativistic spacetime must remain invariant.

The black hole information loss debate to this day fixates on unitarity, a package of symmetries satisfied by deterministic dynamical laws in standard quantum theory. However, the symmetries of unitary laws don't fully coincide with the symmetries of the Einstein Field Equation. It's no wonder then that black hole evaporation appears to be a non-unitary process, in which information turns out to be lost after all.

I'd like to pause and address a repeated objection made by skeptics. They protest that we have insufficient methodological grounds to utilize the states-plus-laws toolkit to begin with (see e.g, Manchak and Weatherall 2018). Not all general relativistic spacetimes are time-orientable and admit of stratification into universal moments (see e.g., Manchak 2011). We simply cannot use the states-plus-laws toolkit in those cases, and the situation for an evaporation spacetime is equivocal at best (Lesourd, 2018). Another overlapping objection is made by scholars steeped in quantum gravity research. They claim we that we risk creating a mess where there isn't one by wielding unfitting tools. Black hole evaporation ultimately falls within the domain of quantum gravity, especially when it shrinks to very small size and emits tremendously high-energy radiation. Many quantum gravity approaches insinuate a pre-spatiotemporal fundamental ontology and will perhaps render the states-plus-laws toolkit obsolete

(see Page 1995; Adlam 2022).

I have two responses to these concerns. My first, more conciliatory response is that these objections may very well be right. However, they would be right for the wrong reasons, which I explore in Section 2.4.1. Refusing to engage further with the majority of the discourse by rejecting an assumption at the outset, whose veracity, I might add, is up in the air, is a wasted opportunity.

Remember that the goal is to definitively establish a paradox, for which I demonstrate the importance of distinguishing between dynamical aspects, having to do with laws, and kinematic aspects, having to do with global states, of black hole information loss. Accomplishing this goal sets up a reductio ad absurdum argument, thus undermining at least one if not several propositions of Hawking's original framework. Should I succeed, I'd be bringing us a major step closer to figuring out whether the spacetime structure and/or the states-plus-laws toolkit have been the culprits all along.

My second, less conciliatory response cautions against throwing the baby out with the bathwater. Physics has benefited immensely from the states-plus-laws toolkit. It carries with it a visceral visualization of the material world as continuously undergoing physical transformations. Evaporation is the epitome of such a process even in the most mundane of human activities, such as boiling water for coffee or tea. It's not surprising that quotidian intuitions about watching boiling water transform into dissipating steam color expectations about black hole evaporation.

Regardless of how the practice of physics evolves for quantum gravity, we should be extremely careful about dismissing the states-plus-laws toolkit from the get-go for black hole evaporation. The mathematical machinery of an initial value problem, where instantaneous data is plugged into partial differential equations, allows us to make concrete predictions. The very enterprise and triumph of modern physics rests on it. Hawking derived black hole evaporation in a semi-classical regime where we want, or dare I say, *need* the states-plus-laws toolkit to function well to make progress, namely to systematize and get a handle on *quantum processes in curved spacetime*, even if that spacetime may well be emergent.

More importantly, even as we've updated and diversified our metaphysical inferences from the states-plus-laws model, we have a strong incentive to keep the states-plus-laws toolkit insofar as we're committed to black hole thermodynamics and statistical mechanics. Come what may vis-à-vis the information loss paradox within Hawking's framework, evaporating black holes in improved frameworks must evolve according to the same laws of thermodynamics as any other familiar evaporating system (see Wallace 2018). Even if quantum gravity has the last word on the statistical mechanical details of black hole evaporation, we had better be able to make the states-plus-laws toolkit work in some effective regime.

2.2.2 Entanglement and Information Conservation

I won't say much more about black hole thermodynamics and statistical mechanics in this chapter, though their contribution to the analysis on information loss is the

import of entropy. Various types of entropy quantify state properties using information measures, the two most important for our discussion being Boltzmann entropy, an information storage capacity measure over degrees of freedom, and von Neumann entropy, a hidden information measure over external entanglement (see Susskind and Lindesay 2004, which is the inspiration for much of the analysis in this section). There's a profound physical connection between entanglement on the one hand, and purity and mixedness on the other hand, all of which are state properties that figure into Boltzmann and von Neumann entropy.

Pure states, whose associated mathematical objects are vectors in a complex Hilbert space, conventionally represent microstates of self-contained systems. A key subtlety to discern, however, is that a self-contained system can exhibit internal entanglement among its parts, represented by a non-factorizable pure state, but the system as a whole is not entangled with its environment. Mixed states, on the other hand, whose associated mathematical objects are density matrices encoding probability distributions over collections of vectors, conventionally represent two categories of inexact states. The first application is for epistemic uncertainty, i.e., statistical ensembles of microstates delineating macrostates. The second application is for what I take to be ontological incompleteness, i.e., entangled subsystems (in line with Susskind et al. 1993; Page 1995; Mermin 1998).

Entanglement is the state property that ties everything together, so let's flesh out its physical interpretation. Presuming realist commitments to quantum states, I advocate the view that entanglement represents a physical property of self-contained systems and a physical relation among subsystems, for two reasons. Bell's theorem proves that entanglement nonlocally fixes correlation structure. 'Locality' has two connotations here. The quantum connotation of locality is that an operation on the state of one subsystem generates correlations with that of another subsystem, which manifestly can't happen in the case of entanglement (see Clifton and Halvorson 2001). Operators representing an entanglement-producing interaction transform a product state into a non-factorizable pure state, and they do so by jointly acting on the pure states of the formerly unentangled subsystems, hence the nonlocality (Clifton and Halvorson, 2001). Realism about quantum states predisposes one to infer that entanglement is a proxy for a holistic physical property of closed systems.

The second, relativistic connotation of locality imposes the speed of light as a nominal upper limit on signaling through a physical substrate (see Maudlin 2002).[1] Since Bell's theorem rules out subluminal signaling among hidden variables, entanglement cannot be produced by dynamically inducing correlations from one subsystem to another, either through a common cause or mediating forces. Nevertheless, the nonlocality of the correlations doesn't preclude subsystems from becoming entangled through interactions that are local in spacetime, which is my reason for characterizing

[1] An alternative reading of relativity imposes the speed of light just as an asymptotic limit on speeds reached by acceleration/deceleration, where faster speeds are theoretically allowed (see e.g., Asaro 1996). However, Maudlin (2002) argues that the direction of time forbids superluminal signaling due to backwards influence.

entanglement as a physical relation. As I see it, entanglement is a property of the self-contained system as well as the parts, and consequently, a property of the relation as well as the relata.

The entanglement-based incarnation of von Neumann entropy, S_{VN}, quantifies the extent of hidden information stored in nonlocal correlations with the environment, i.e., correlations with missing degrees of freedom. It's a logarithmic measure of a probability distribution (where $0 \leq p_i \leq 1$ and $\sum p_i = 1$) over pure states, the ensemble of which is represented as a density matrix ρ (refer to Equation 2.1):

$$S_{VN} = -tr\rho \ln \rho = \sum_{i=1}^{n} p_i \ln p_i;\ p_i \geq 0,\ \Sigma_i p_i = 1. \tag{2.1}$$

I conceptualize the probability distribution in this context as representing ontological incompleteness because ρ is the reduced density matrix of a subsystem obtained by tracing out degrees of freedom of the entangled complement.

Von Neumann entropy has a twin statistical incarnation as well, which I prefer to distinguish as Gibbs entropy, S_G, where the probability distribution represents epistemic uncertainty over pure states in an ensemble delineating a macrostate. Both information measures share the same mathematical form, but they do not conceptually reduce to each other for many quantum state realists (see e.g., Susskind et al. 1993).

However, one may counter that the probability distribution in von Neumann entropy also implies epistemic uncertainty, particularly uncertainty about the state of the subsystem upon projective measurement, after which entanglement is severed (see Earman 2015). My initial, tangential rebuttal is that projective measurement counts as a local, non-unitary dynamical law (Clifton and Halvorson, 2001), which is irrelevant to a puzzle about global non-unitary evolution without any projective measurements taking place on the entire universe, that is, presuming the absence of an omnipotent superobserver periodically interrupting the peace.

More importantly though, prior to collapse, the mixed state of an entangled subsystem reveals that it's in a unique state, often a superposition of states, but the uncertainty revolves around the nonlocal correlation with the traced out degrees of freedom, not the subsystem at hand (see e.g., Page 1995). Pure states, on the other hand, have no such hidden information stored in nonlocal correlations with external subsystems; they're descriptively exhaustive and contain information about maximum Boltzmann entropy, S_B, a logarithmic measure of Hilbert space dimensionality, $dim(H)$ (refer to Equation 2.2):

$$S_B = \ln(dimH). \tag{2.2}$$

Any global state should undeniably maximize the entire universe's Boltzmann entropy, considering that there can't be any missing degrees of freedom. The values of all degrees of freedom must be specified to locate a unique unit vector in Hilbert space, which is why the representation of global states with pure states is natural. Likewise, any global state should also minimize von Neumann entropy; after all, there's no physical environment beyond the entire universe to store nonlocal correlations. Another apt feature of pure states is that their von Neumann entropy vanishes. As such,

pure states are taken to always denote self-contained systems without any external entanglement.

Mixed states of entangled subsystems, on the flip side, always have positive von Neumann entropy, seeing as realism about entanglement insinuates ontological incompleteness. In fact, any mixed state with positive von Neumann entropy must be derived from a pure state with zero von Neumann entropy, which Calosi and Morganti (2021) argue signals ontological dependence.

An entangled subsystem's maximum Boltzmann entropy bounds its von Neumann entropy from above because it's unintelligible to have more entanglement than degrees of freedom available to be entangled. Indeed, von Neumann entropy is maximized when it's equal to Boltzmann entropy, which happens for mixed states with uniform probability distributions (see Susskind and Lindesay 2004). Analogously to the situation with von Neumann entropy, maximum Boltzmann entropy bounds Gibbs entropy from above, and they converge for uniform probability distributions. Unlike mixed states with positive von Neumann entropy, however, mixed states solely with positive Gibbs entropy do not depend on anything else, ontologically or otherwise.

The salience of all these information measures, with particular focus on Boltzmann and von Neumann entropy for the remainder of the analysis, is that they're conserved under unitary evolution (Susskind and Lindesay, 2004)). Consider the variety of input states that a unitary operator can act on: pure global state, pure state of self-contained subsystem, mixed state of entangled subsystem, mixed state of global macrostate, mixed state of subsystem's macrostate, etc. Unitary operators ensure that the output state is confined to the same category: like begets like. Intuitively speaking, unitary operators fix the referent system during the evolution from the input to output state. This fixing is quite convenient because unitarity maintains information storage capacity and by extension, it also preserves external entanglement in the absence of further interactions, an implication that has gone unappreciated in the black hole information loss discourse.

2.3 Foreshadowing a Paradox

It should hopefully be straightforward to foresee that non-unitarity threatens the conservation of Boltzmann and von Neumann entropy. As I explained in Chapter 1, there are four stacked levels at which unitarity can be violated, each tied to its own information conservation principle. Black hole evaporation shatters the foundation. Neither Boltzmann nor von Neumann entropy is conserved. The former decreases, signifying the elimination of degrees of freedom, and the latter increases, intimating the appearance of external entanglement. In this section, I introduce a toy model, that of Hawking pair production, to ground information loss in concrete dynamical mechanisms and set the stage for the paradox of phantom entanglement.

2.3.1 Overview of the Black Hole Information Loss Puzzle

To summarize the black hole information loss puzzle, global dynamical evolution involves an uncanny pure-to-mixed transition. Pre-evaporation global states are pure, representing how the black hole interior and exterior are entangled. Post-evaporation global states, however, are mixed, representing Hawking radiation after the disappearance of the event horizon. The situation wouldn't be so dire if Hawking had calculated proper mixtures of global macrostates. But the density matrices he produced involve improper mixtures, where the physical interpretation of the probability distributions is underdetermined between epistemic uncertainty of statistical ensembles and ontological incompleteness of entanglement.

As I argued in Chapter 1, we know that if a black hole evaporates completely, late-time Hawking radiation can't be a subsystem. It only appears to be so because degrees of freedom were in fact eliminated, which is reflected in the decrease in maximal Boltzmann entropy. And throughout evaporation, Hawking radiation had been entangled with negative-energy modes headed towards the singularity. So, Hawking radiation may still be entangled, as reflected in the increase in global von Neumann entropy. We're in the bizarre circumstance of handling an externally entangled global system, as physicists like Bekenstein (1994), Mathur (2009) and Mann (2015) confess but swiftly sweep under the rug.

In order to decisively deem black hole information loss paradoxical, I must demonstrate a clear-cut contradiction inherent to the idea of an externally entangled global system. But because we're trying to interpret a density matrix for a novel physical situation, we risk wires getting crossed. To dispel confusion, I've refined a toy model of Hawking pair production, one of the proposed dynamical mechanisms for Hawking radiation. I review the physics behind Hawking pair production in the next section, and then I round off the analysis in the subsequent section by illustrating the process in a Penrose diagram of an evaporation spacetime. This toy model is the ladder to the paradox of phantom entanglement, which will eventually be kicked away.

2.3.2 Hawking Pair Production in a Semi-Classical Framework

The quantum vacuum is infamous for being much more interesting than its classical counterpart. The goal is to solve the Klein-Gordon wave equation of free bosonic scalar fields such that the Lorentz-invariant field values are zero everywhere. In coordinate systems anchored to stationary observers outside the event horizon, solutions can be decomposed into destructively interfering waves of positive and negative-frequency modes that propagate infinitely far away and into the black hole respectively (Hawking, 1975).

A heuristic but physically intuitive picture for the quanta of these waves is that of localized positive and negative-energy field excitations or particles, depending on one's preferred ontology. This physical picture seems to invalidate the notion of a vacuum

without particles or energy whatsoever, but the idea is that they form anti-particle pairs that could potentially annihilate, recovering an expectation value of net zero particle number. Members of these Hawking pairs are also entangled to settle the energy debt.

Since Hawking pairs are perfectly balanced, they can be split into complementary entangled subsystems of positive and negative-energy quanta respectively. In fact, members of Hawking pairs are "monogamously entangled", whereby they saturate their mutual entanglement. Therefore, the von Neuman entropy of positive and negative-energy subsystems is maximized to equal their Boltzmann entropy, which implies density matrices of uniform probability distributions (for discussions on Hawking pairs modeled as qubits of Bell states, see Mathur 2009; Susskind 2012; Mann 2015; Osuga and Page 2018).

The quantum vacuum is teeming with so-called virtual activity, so what started out as a vacuum state doesn't end up as one. The event horizon, the global boundary from which not even light can escape, separates members of Hawking pairs. Negative-energy quanta that tunnel into the black hole – a classically-forbidden trajectory because gravity is attractive only when mass/energy is positive – plummet towards the ill-fated singularity, while their positive-energy partners escape as radiation.[2]

Since the speed-of-light barrier prohibits members of Hawking pairs from recombining, a stationary observer outside the event horizon can indeed detect the surviving positive-energy quanta, whose frequency distribution obeys the Planck spectrum of black body radiation (Hawking 1975; Hawking 1976).

> A stationary observer outside the black hole could regard a particle he detected... as being one member of a pair of particles created by the gravitational yield [sic] of the collapse, the other member having negative energy and having fallen into the black hole (Hawking, 1976, p. 2468).

An infalling observer, however, would not, for all intents and purposes, detect Hawking radiation. The distribution of field modes as determined in coordinate systems of locally inertial reference frames is continuous across the event horizon, not split evenly into positive and negative frequencies. In this alternative decomposition, negative-frequency contributions are much smaller, so the relative proportion of Hawking pairs with members separated across the event horizon is negligible (Hawking, 1975).

But in order for that proportion to stay negligible, the local radius of curvature must be larger than the Planck length of 10^{-33} cm. The smaller the radius of curvature, the larger the local energy density. Not only would we need a theory of quantum gravity to take over at Planck scales, large local energy densities increase the amount of detectable Hawking radiation for infalling observers. However, because the event horizon is a global boundary, no local phenomena should mark its location. For that reason, Hawking (1975) avers that particle creation is a global process. In sub-Planckian regimes at least, the experience of a minimally-disturbed, adiabatic vacuum

[2] Parikh and Wilzcek (2000) formalize the derivation of Hawking radiation as a dynamical process of members of particle-antiparticle pairs tunneling across the horizon.

for infalling observers recovers the general relativistic prediction of what Almheiri et al. (2013) have popularized a "drama-free horizon".

2.3.3 Non-globally Hyperbolic Spacetime Structure

Given that Hawking pair production is sensitive to global spacetime structure, it makes sense to take a closer look at an evaporation spacetime, paying particularly close attention to global features such as the event horizon, singularity, and topology. This process is illustrated in Figure 2.1, a more detailed version of Hawking's own Penrose diagram (see Hawking 1975) featuring an evaporating Schwarzschild (i.e., non-rotating and uncharged) black hole.

The spacetime structure and foliation scheme of Figure 2.1 have consistently been reproduced in the literature as the springboard for black hole information loss (see Susskind et al. 1993; Belot et al. 1999; Chen et al. 2015; Lesourd 2018). I will not entertain alternative evaporation diagrams or foliation schemes for the purpose of setting up a robust paradox as I wish to refine the mainstream narrative and proffer a more sophisticated account of black hole information loss while holding fixed widely-held baseline assumptions, with an eye towards expounding quantum gravitational proposals of black hole evaporation and black hole thermodynamics/statistical mechanics down the line. In order to facilitate our understanding of the conventional concerns over black hole evaporation, I've further annotated it to display heuristic visualizations of the quantum matter fields.[3]

Like a conventional Penrose diagram, the vertical axis is the global time parameter ranging from past to future timelike infinity, i^- to i^+. The horizontal axis is the radial parameter from the center of the black hole, ranging from $r = 0$ to spacelike infinity, i°. Penrose diagrams preserve conformal structure (the fact that for $c = 1$, light rays travel at 45° angles), but they don't preserve metrical structure, which is readily apparent since infinite distances are squashed to finite distances. The event horizon is represented by the 45° line shrouding the singularity to indicate a lightlike (null) boundary.

Unlike a conventional Penrose diagram, however, an evaporation spacetime continues beyond the singularity. It's convenient to treat E as a point on the vertical, spacetime axis at $r = 0$ as opposed to a point on the horizontal, singular segment to avoid the pathologies of naked singularities (Hawking, 1975). That way, we can interpret E as the final evaporation event marking the disappearance of the event horizon in spacetime (Maudlin, 2017).

Moreover, Hawking's simplest derivation involves non-interacting quanta of a massless scalar field with 45° lightlike trajectories. They all originate at past null infinity, $\mathcal{I}^-$. The quanta that end up at future null infinity $\mathcal{I}^+$ have positive energy (in red), whereas the quanta that are terminated at the singularity have negative energy (in blue). Strictly speaking, these quanta don't consistently have positive and negative-

[3] I couldn't have produced any of the following Penrose diagrams without Baptiste Le Bihan's contributions.

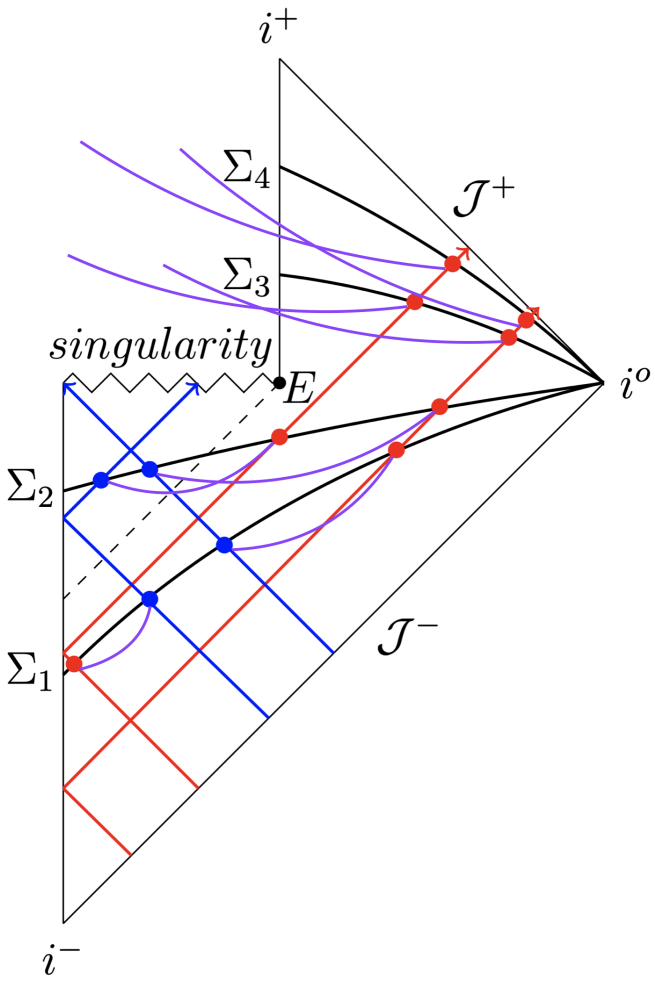

Figure 2.1: Evaporating Black Hole

Red and blue nodes display where positive and negative-energy trajectories (respectively) intersect four spacelike hypersurfaces, Σ_1, Σ_2, Σ_3, and Σ_4. Purple curves portray entanglement between members of Hawking pairs. Σ_3 and Σ_4 only contain red nodes, indicating that Hawking radiation consists of solitary positive-energy quanta. It's an open question whether they're still entangled.

energy throughout the entire spacetime. Instead, their designation depends on if they're ingoing or outgoing with respect to different regions, such as $\mathcal{I}^-$ or $\mathcal{I}^+$. Furthermore, a subset of quanta only acquire negative-energy upon crossing the event horizon. However, this is a heuristic visualization, and the important takeaway is to keep track of entangled Hawking pairs as if they traversed the entire spacetime.

Monitoring entangled Hawking pairs is facilitated through intersections between the maximally extended worldlines of massless quanta and continuous spacelike curves like Σ_1, Σ_2, Σ_3, and Σ_4. Portrayed as nodes, these intersections are proxies for matter degrees of freedom figuring into a Hilbert space representation. Matter degrees of freedom follow causal trajectories of subluminal or luminal speeds, indicated by their wordlines having tangent slopes greater than or equal to 45° (corresponding to timelike and lightlike separation respectively). These moving systems can always be

transformed to proper rest frames in which the spatial separation is reduced to zero.

On the flip side, the tangent slope at every point on a continuous spacelike curve is less than 45°, indicating distances in which it's the temporal separation that can be reduced to zero. That's why Σ_1, Σ_2, Σ_3, and Σ_4 represent spatial hypersurfaces and are utilized as simultaneity slices, registering the global ontology at any given moment with blue and red nodes. Finally, only spacelike separated blue and red nodes intersecting the same spatial hypersurface can be connected by purple curves depicting entanglement. That's because in order to make sense of black hole evaporation in terms of temporal evolution, quantum states are defined on simultaneity slices. However, the subset of red nodes with straggling purple entanglement curves due to the absence of spacelike separated blue nodes at that instant are the stars of the show and will be contemplated thoroughly in Section 2.5.

To draw potentially contradictory conclusions, we need some basic technical machinery from general relativity and quantum field theory. A Cauchy surface is a special type of spatial hypersurface that contains all possible nodes. In this diagram, a Cauchy surface can intersect at most four lightlike trajectories, so it likewise contains at most four nodes. The agglomeration of all such nodes constitutes "Cauchy data" (see (Hawking, 1975)). For a vacuum state, Cauchy data consists of an equal number of blue and red nodes (two each) connected by purple curves. The physical explanation is that entangled Hawking pairs have expectation values of zero particle number and energy.

Notice the finite count of degrees of freedom, which undoubtedly stems in part from practical considerations. To be fair, all hypersurfaces extending to spacelike infinity i^0 possess infinite volume and technically contain countably infinite degrees of freedom, which is expected for quantum scalar fields modeled in Fock space (see Ruetsche 2011). All countably infinite sets also have the same cardinality, making the business of comparing quantities and information measures, like Boltzmann and von Neumann entropy, across states controversial.

Though recent progress has been made in defining entanglement entropy in the large-N limit (see Chandrasekaran et al. 2023). we don't care about all degrees of freedom in this toy model. We're solely concerned with those involved in Hawking pair production that occupy bounded spatial regions. As per Christodoulou and Rovelli (2015), the black hole interior is a bounded spatial region with large but finite volume, the magnitude depending on the choice and value of the time coordinate fixing the spacelike hypersurface. Bounded spatial regions in this quantization procedure always have finite particle number and energy densities (Ruetsche, 2011).

Therefore, interior degrees of freedom must be associated with finite Boltzmann and von Neumann entropy. Since Hawking pair production entails two complementary entangled subsystems of equal Boltzmann and von Neumann entropy, the same holds for exterior degrees of freedom. Without loss of generality then, we can straightforwardly quantify and compare the desired matter degrees of freedom across states.

Let's proceed to evaluate Figure 2.1. Σ_1 is a simultaneity slice prior to the formation of the black hole, which is evident because it does not cross the event horizon. It

contains two sets of entangled Hawking pairs with an equal number of red and blue nodes connected by purple curves. Therefore, it's both a Cauchy surface and a vacuum state, expressed as a pure global state maximizing Boltzmann entropy and minimizing von Neumann entropy.

The Cauchy data of the subsequent hypersurface, Σ_2, mirrors that of Σ_1, but its unique spatiotemporal location captures the production of Hawking radiation. By this time, the event horizon permanently divides the positive-energy quanta (red nodes) from their negative-energy partners (blue nodes), with entanglement between the black hole exterior and interior.

Next, consider Σ_3, a post-evaporation simultaneity slice. It contains only two red nodes; therefore, it's neither a Cauchy surface nor a vacuum state. It appears that the red nodes are all that's left of the global system, hence the reduction in Boltzmann entropy, and it's these solitary positive-energy quanta that constitute late-time Hawking radiation. Strangely enough, purple curves are still present, hinting at lingering entanglement and positive von Neumann entropy. Σ_4 tells the same story.

It should be apparent now that Σ_3 and Σ_4 are those perplexing mixed radiation states upending standard quantum theory. Do they reflect global states with vanishing von Neumann entropy or do they intimate that positive-energy quanta are still entangled with their negative-energy partners? The pure-to-mixed transition from Σ_1 to Σ_4 is a new flavor of non-unitarity – unrelated to measurement collapse – that cries out for interpretation.

A promising explanation is that the evolution is Cauchy-to-non-Cauchy. As Manchak and Weatherall (2018) put it, "[G]lobal hyperbolicity appears to be necessary for global unitary evolution in the context of quantum theory set in curved spacetimes" (p. 615).[4] Hawking's derivation demonstrates that an evaporation spacetime is not exhaustively foliable by Cauchy surfaces, and therefore, not globally hyperbolic, so exclusively pure-to-pure evolution seems doomed to fail.[5]

What I hope to parlay by the end of this chapter is that while non-global hyperbolicity accounts for much for the mystery, it doesn't account for all of it. Figure 2.1 foreshadows my formulation of the paradox of phantom entanglement. The purple curves connecting red nodes on one end but left hanging on the other end exhibit how late-time Hawking radiation is entangled with unphysical degrees of freedom. I ar-

[4]It surprised and confused me when Manchak and Weatherall (2018) assert on the same page that global hyperbolicity is sufficient but not necessary for global determinism. Given that global unitarity ensures global determinism in any quantum theory, holding on to the necessity of global hyperbolicity for global unitarity entails a contradiction. I suspect – based on the subsequent discussion in Chapter 3 – that global hyperbolicity is indeed sufficient but not necessary for global unitarity. However, insofar as we're restricting attention to conservative spacetimes, where dynamical influences propagate along timelike or lightlike trajectories in the absence of closed causal curves, global hyperbolicity is both sufficient and necessary for global unitarity.

[5]Theorems in general relativity forbid continuous spacetimes from admitting only some Cauchy surfaces – it's an all or nothing deal. However, evaporation spacetimes are topologically discontinuous, thus opening the door to a hybrid stratification structure. Regardless, relativistic spacetimes undergoing topology change undermine global hyperbolicity (see Belot et al. 1999).

gue that Hawking's semi-classical formalism accommodates viewing the final radiation state as an entangled subsystem, despite the lack of guidance about how to locate the complementary entangled subsystem, an issue I will flesh out in Section

Before I move on, I'd like to acknowledge the elephant in the room. You may be wondering whether the stray, purple entanglement curves are merely an artifact of a poor choice of global simultaneity slices. The red nodes above E are spacelike separated from the black hole interior because they're outside the future lightcone of the event horizon, as illustrated in Figure 2.2. In other words, they fall within the domain of dependence Σ_1, under the pseudo-Cauchy horizon in yellow, allowing for continuous Cauchy surfaces to slice across red and blue trajectories prior to E. This inference is bolstered by the spacelike nature of the purple entanglement curves linking the red nodes on Σ_3 with blue nodes inside the black hole.[6]

Thus, an alternative foliation scheme would reveal that the black hole does not evaporate prior to Hawking radiation reaching future null infinity; the two entangled subsystems exist concurrently, solving the mystery behind radiation states always being mixed. Σ_3 and Σ_4 could be disregarded and all Hawking pairs could be encompassed in a globally hyperbolic region embedded in the non-globally hyperbolic spacetime. It's the top-left region of the evaporation diagram in the future lightcone of the final evaporation event that prompts non-global hyperbolicity, since it doesn't admit of simultaneity slices spanning from $r = 0$ to i^0. That would nonetheless be inconsequential as no Hawking particles following null trajectories would make it out there. The kicker of the pseudo-Cauchy horizon is that black hole evaporation within a semi-classical framework could be unitary for all intents and purposes, an insight I've not come across in any publications.

While I concede this technical loophole, the mainstream narrative decrying non-unitarity has been sidelined it as a viable possibility, so I briefly address in Section 2.5 and Section 2.6 as to why it may have been perceived by some as a hollow victory. In Section 2.5, I additionally brainstorm scenarios in which we would be compelled to care about the top-left region of the evaporation diagram to describe Hawking radiation, where the positive-energy survivors would indeed be left with stray, purple entanglement curves.

Regardless of the outcome of such a project, which I will not be embarking on, demonstrating a foliation-dependent paradox nevertheless suffices for my dialectical aims. The mainstream narrative anchoring its conception of black hole information loss to the evaporation diagram in tandem with the foliation scheme of Figure 2.1 has been vacillating on the status of a paradox for the wrong reasons, considering that non-unitarity simpliciter doesn't entail a contradiction. By setting up a more rigorous paradox that the mainstream narrative should've been worried about *at a particular stage in the discourse*, and in fact was implicitly worried about, I'm able to sharply pose a universal question about black holes that spares nobody – including those opting for the technical loopholes.

[6] I'm beyond grateful to Dominic Ryder for his patience in helping me understand this nuance.

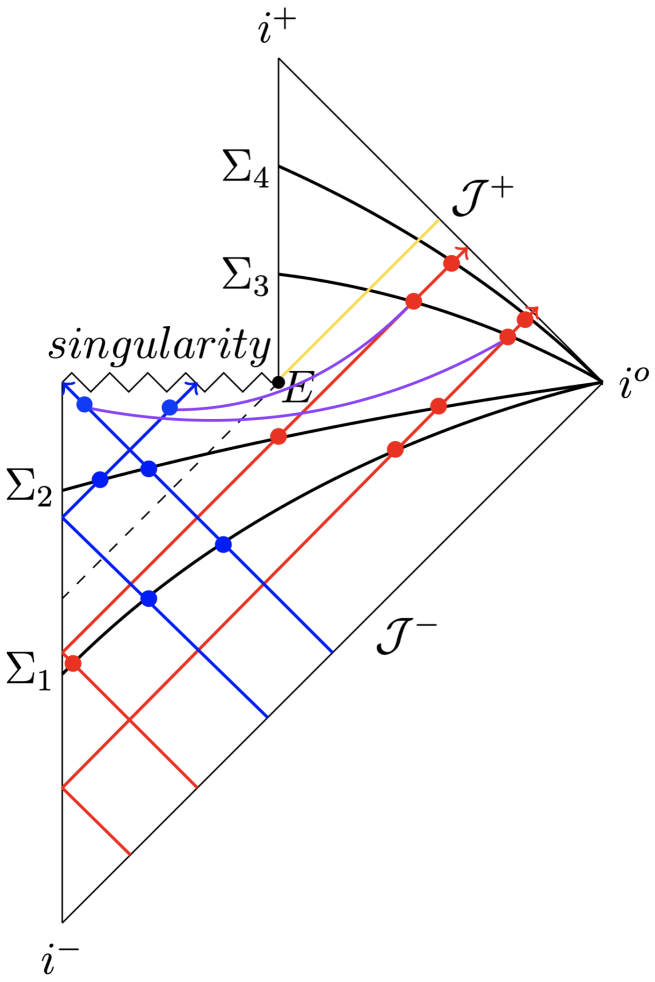

Figure 2.2: Technical Loophole for Unitary Evaporation
The yellow line extending the dashed event horizon through E demarcates the future lightcone of the final evaporation event. The spacetime region below – containing the red trajectories of positive-energy quanta – falls within the domain of dependence of Σ_1 and is globally hyperbolic. Consequently, it's possible for the red nodes on Σ_3 to be entangled with blue nodes in the black hole interior (depicted with spacelike purple curves) as if they intersected a continuous Cauchy surface in between Σ_2 and E. The same holds for the red nodes on Σ_4 (not illustrated).

2.4 Competing Dynamical Narratives: General Relativity versus Quantum Theory

In this section, I distill and meticulously spell out how the physical interpretation of mixed Hawking radiation states like Σ_3 and Σ_4 hinges on a two-fold dynamical narrative: 1) the general relativistic argument and 2) the quantum theoretic argument. Each argument emphasizes its respective theory's contributions to the overall, semi-classical dynamical evolution, individually reaching a prima facie acceptable conclusion.

The diverging opinions over the paradoxical status of the black hole information loss puzzle can be traced to the disproportional treatment each argument has received

in physics and philosophy discourse. The general relativistic argument has engendered greater controversy, which I argue is separately defeasible. The quantum theoretic argument, in contrast, has gone virtually unnoticed in the literature and is in fact the potent ingredient in the impending paradox of phantom entanglement – cooked up with the conjunction of both arguments' conclusions.

To reiterate, however, the paradox of phantom entanglement is not supposed to be an airtight, inescapable paradox as per today's standards. Both the general relativistic and quantum theoretic arguments employ loaded premises susceptible to objections – leaving more options for purported solutions than I'd care to count. Otherwise, semi-classical gravity wouldn't be a precarious theory in and of itself nor a stepping stone to quantum gravity. My goal is to argue that any self-proclaimed solution deserves its designation not because it's naively restoring unitarity, but rather, because it's upending a legitimate paradox given the baseline assumptions under which it was initially operating. My job is to elucidate the original premises vulnerable to refutation. Most importantly, retroactively reconstructing the paradox of phantom entanglement allows me to clean up the discourse and organize *all* proposals on the table based on a powerful guiding principle bearing on black hole thermodynamics and statistical mechanics.

2.4.1 The General Relativistic Argument

I mentioned in Section 2.3 that a promising explanation for non-unitary evolution in the black hole information loss puzzle is non-global hyperbolicity. To get a sense of why, let me return to another statement I made in passing, that global hyperbolicity may be a necessary condition for global unitarity.

As a reminder, globally hyperbolic spacetimes are exhaustively foliable by Cauchy surfaces, which intersect all inextensible worldlines exactly once (see Malament 2012). So, they're a restricted class of general relativistic spacetimes in which a succession of Cauchy surfaces strings together the universe's unique history. For that reason, transitions between Cauchy surfaces, e.g., between Σ_1 and Σ_2, conserve total degrees of freedom and maximal Boltzmann entropy, both of which are part and parcel of unitarity.

Transitions between non-Cauchy surfaces may also conserve Boltzmann entropy if their sets of degrees of freedom match in cardinality (see Susskind and Lindesay 2004). For example, Σ_3 and Σ_4 have the same quantities of degrees of freedom, and the transition is unitary precisely because it's mixed-to-mixed.

We cursorily saw in the previous section that transitions between Cauchy and non-Cauchy surfaces, however, conserve neither degrees of freedom nor Boltzmann entropy. Yet, we did not isolate the cause behind there being Cauchy-to-non-Cauchy transitions in the first place, which is tied to non-global hyperbolicity.

What I want to accomplish in the general relativistic argument is to highlight the glossed over connection between non-global hyperbolicity in evaporation spacetimes and a variation in degrees of freedom. Thus far, we've learned a couple of important

things. If we split an evaporation spacetime into two regions, one below and another above the final evaporation event, E, each region taken as a separate spacetime is globally hyperbolic and hospitable to unitarity.

We can also concretely see that by conserving degrees of freedom, global hyperbolicity and unitarity preserve the amount of the ontology that's registered on a state over the course of dynamical evolution. So, it appears that the infamous pure-to-mixed transition occurs as a momentary blip at the final evaporation event, E. The culprit uprooting ontological stability is a topological discontinuity.

The topological discontinuity at E marks the detachment of the black hole interior from the rest of the spacetime, resulting in the failure of global hyperbolicity. The hybrid stratification structure indicates that negative-energy (blue) trajectories cannot reach and intersect a hypersurface like Σ_3; the event horizon is a causal barrier and the singularity cuts a region of spacetime short. A direct ramification of this is that worldlines registering on Σ_1 and Σ_2 but failing to register on Σ_3 and Σ_4 belong to missing ontology along the global time parameter.

The evolution from Σ_2 to Σ_3 involves an elimination of degrees of freedom precisely because the latter isn't a Cauchy surface, and therefore, contains a proper subset of degrees of freedom compared to prior Cauchy surfaces. Blue trajectories truly depict lost, negative-energy degrees of freedom in the topologically disconnected black hole interior. The dynamical mechanism accounting for the change in maximal Boltzmann entropy is the singularity that terminates worldlines.

On the face of it, it's not too surprising that unitarity malfunctions – period – in evaporation spacetimes. Seeing as the initial and final states involve different sets of degrees of freedom, unitarity must fail on pain of consistency. With annihilation on the table, the most readily available interpretation of the states defined on Σ_3 and Σ_4 is that of global mixed states capturing the remaining ontology. The decrease in maximal Boltzmann entropy literally translates into a reduction of global information storage capacity. It's unclear whether late-time Hawking radiation is fundamentally in a mixed state or is associated with some unknown pure state of fewer degrees of freedom. Either way, Σ_3 and Σ_4 must be mixed with respect to the original Hilbert space.

Before I outline the general relativistic argument, I wish to address a couple of potential sources of confusion. On the face of it, one may wonder what's so radical about the annihilation of negative-energy degrees of freedom. The guarded suggestion of Hawking (1976) that positive and negative-energy quanta pop into existence out of fluctuations and restore the vacuum state through self-annihilating collisions[7] – perpetuating a heuristic understanding of creation and annihilation operators – actually makes the observation about non-global hyperbolicity somewhat obsolete.

Fluctuations presume non-unitary dynamics. Eigenstates of energy and occupation

[7]Creation and annihilation operators do not represent physical observables in standard quantum theory. Only self-adjoint operators conventionally represent physical observables, whereas creation and annihilation operators are each other's adjoints; however, they can be combined into self-adjoint operators, notably corresponding to the physical observable of occupation number (Ruetsche, 2011).

number are stationary under unitary evolution, whereas a fluctuation entails jumping from one eigenvalue to another, which is exactly what projective measurement or spontaneous collapse is designed to do. So, if we were to take vacuum fluctuations seriously for the production of Hawking radiation, the dynamics would be non-unitary irrespective of the black hole singularity.

Expecting global hyperbolicity would then be utterly incoherent. How could there be Cauchy surfaces keeping track of the global ontology when worldlines could begin and end anywhere and fail to span the entirety of the spacetime? Non-unitarity from repeated fluctuations would thus precede and generate non-global hyperbolicity, in which case the entire discourse over information loss would be moot – this is the position of Okon and Sudarsky (2017). However, the direction of explanation is inverted and the explanation too generic. It would be true of any spacetime, evaporation or not, with quantum states of matter fields undergoing collapse.

Moreover, the production of Hawking radiation is not contingent on a literal interpretation of vacuum fluctuations. As Belot et al. (1999) clarify, black hole evaporation is non-unitary even without a collapse mechanism. The strength of Figure 2.1 is portraying how the most conservative dynamical situation – with non-interacting positive and negative-energy quanta that have been hanging around since past null infinity and evolving unitarily whenever conditions allow – nevertheless winds up non-unitary. Non-global hyperbolicity, arising from the singularity and topological discontinuity, precedes and generates a short-lived, non-unitary transition through E of huge impact. Degrees of freedom associated with the physical observables of an entire spatial region, the black hole interior, are eliminated. Here, the direction of explanation accounts for the uniqueness of black hole evaporation.

Another potential source of confusion may be the distinction between evaporating black holes and their classical counterparts in possessing such destructive capabilities. After all, classical black holes also shroud singularities that terminate the trajectories of entities trapped behind the event horizon. The difference is subtle, yet crucial. In the classical case, like the globally hyperbolic Schwarzschild solution, not all observers agree that infalling systems have their existence ended prematurely.

In their own reference frame, infalling systems certainly confront the singularity after finite proper time (see Earman 1996). But in reference frames parameterized by the coordinate time of a global foliation, observers disagree that infalling systems confront the singularity after finite coordinate time. No matter how much coordinate time passes, a late-time Cauchy surface intersects all interior and exterior trajectories because the singularity and future timelike infinity are spacelike separated. As such, a classical black hole traps interior degrees of freedom indefinitely due to the causal barrier posed by the event horizon, but its shrouded singularity doesn't outright annihilate them (see e.g., Maudlin 2017). To understand better, it's helpful to review the meaning of 'singularity'.

What different types of spacetime singularities, such as curvature blowups or missing parts of the manifold, have in common is geodesic incompleteness. Geodesic incompleteness implies that a subset of worldlines, like those entering a black hole, are

extended just finitely into the future, as opposed to infinitely. Yet a singularity itself is *outside of spacetime* (Earman, 1996). I take 'annihilation', however, to connote the elimination of an entity *in spacetime*. 'Annihilation' also connotes there being measurable temporal intervals following the disappearance of an entity. It's patently absurd to refer to one's own annihilation, suggesting that it's an inappropriate concept to apply to inertial observers following incomplete geodesics.[8]

Moreover, classically there exists a physically reasonable global reference frame in which infalling systems should also not be considered annihilated because it actually takes infinite coordinate time for them to confront the singularity, after which there is similarly no measurable temporal interval. As Maudlin (2017) reminds us, not all points on a Penrose diagram belong to spacetime. Because the singularity and future timelike infinity are open edges of spacetime, not spacetime points, their spacelike separation can superficially be understood as delineating the edge of time.

The game changer for black hole evaporation, however, is the disappearance of the event horizon and the extension of spacetime beyond the final evaporation event. Because the singularity and future timelike infinity are no longer spacelike separated as revealed by the conformal structure of an evaporation Penrose diagram (see Figure 2.1), any global foliation of the spacetime admits exclusively of observers for whom infalling systems do indeed confront the singularity in finite coordinate time.[9] At some moment like Σ_3 (which we've already taken for granted is a simultaneity slice based on motivations provided in Section 2.2.1), the fact that no worldlines having entered the black hole are present indicates that they got terminated prematurely by the singularity in the past, and the distance to the past is calculable by taking the final evaporation event, a proper spacetime point, as the start of the temporal interval. All observers now agree that infalling systems have their existence cut short.

Figure 2.1 strikingly captures how quantum field theory on a Schwarzschild geometry produces Hawking radiation – surviving positive-energy quanta outright lose their negative-energy partners in a region that becomes causally cut off, which motivates why the black hole's mass reduces in a perfectly counterbalancing act should back-reaction be taken into account. As such, the general relativistic argument concerns a variation in degrees of freedom from pre-to-post-evaporation, where I endeavor to show that Σ_3, which contains fewer matter degrees of freedom than Σ_2, is nevertheless translatable as a global state.[10]

The general relativistic argument:

1. All degrees of freedom of Σ_2, a pre-evaporation Cauchy surface, constitute the global pre-evaporation system.

[8] I'd like to thank Baptiste Le Bihan for input about the nuances of 'annihilation'.

[9] Maudlin (2017) proposes foliating Hawking's Penrose diagram differently to reintroduce observers for whom infalling systems confront the singularity in infinite coordinate time, like in the classical case. Given pushback against his view, I'm unsure what to make of it but will not engage further here.

[10] For an extended version of this argument, see Appendix A.

2. A proper subset of Σ_2 degrees of freedom are conserved on Σ_3, a post-evaporation non-Cauchy surface.

3. A proper subset of Σ_2 degrees of freedom are annihilated by the black hole singularity.

4. Only conserved degrees of freedom constitute the global post-evaporation system.

Conclusion:

5. Therefore, the degrees of freedom of Σ_3 constitute the global post-evaporation system.

As anticipated of any physical system that evaporates completely, solely radiation remains. That's what happens when water turns into vapor, or when stars burn away their energy. However, black holes are critically different, and the general relativistic argument illuminates how reaching a conclusion resembling the evaporation process of other physical systems is nontrivial.

In globally hyperbolic spacetimes, the system that's spatially global at an instant of time is always associated with a Cauchy surface at every moment over the course of its evolution. For example, in a universe with dying stars but no black holes, the global post-evaporation system comprised only of radiation is associated with a Cauchy surface. All degrees of freedom are conserved throughout the universe's history, and evaporation is but another physical process of transformation.

However, black hole evaporation is incompatible with global hyperbolicity as per Hawking's calculation. The steps of the general relativistic argument motivate why simultaneity slices like Σ_3, which aren't Cauchy surfaces, nonetheless correspond to the global post-evaporation system comprised only of Hawking radiation. Far from being merely a physical process of transformation, black hole evaporation removes world-building blocks, their properties, or both. Under these circumstances, is the evolution of Hawking radiation from a subsystem into a global system due to the loss of black hole interior degrees of freedom a prima facie acceptable conclusion? Bokulich (2011) contemplates what would lead one to answer with a resounding yes:

> [I]f one takes the space-time geometry of general relativity seriously, not only might one accept the loss of information, one might insist that it is an inevitable and unproblematic consequence of the existence of black holes (p. 372).

Despite the appearance of a romanticized clash between relativists and particle physicists, relativists disagree among themselves about the extent to which we can trust the geometry of an evaporation spacetime, which doesn't utilize a differentiable manifold like a standard relativistic spacetime. I'd like to address a dispute among philosophers who are sympathetic to various relativists' attitudes towards black hole information loss, that is Maudlin (2017) on the one hand and Manchak and Weatherall

(2018) on the other hand, about the physical reasonableness of Hawking's Penrose diagram.

The premise under scrutiny is the third one, a compressed summary of the repercussions of global non-hyperbolicity. What's contentious about its contents – besides the urge to jump the gun and rely on singularity resolution in quantum gravity – is the physical reasonableness of the topological discontinuity at the final evaporation event, E. There the metric is undefined, yet it exerts influence on the physics everywhere else. Maudlin (2017) is unfazed by E; he reasons that the final evaporation event can heuristically be thought of as a spacetime point glued to the rest of spacetime. He asserts that its lightcone (i.e., causal) structure is consistent with that of all other spacetime points, although physicists face the non-trivial task of defining the metric by hand at E and specifying the novel geometry and physical laws that obtain there. Nevertheless, his position is that the spacetime structure of Hawking's Penrose diagram is perfectly well-behaved and compatible with tame dynamics throughout black hole evaporation.[11]

> If we adopt this account of the space-time then we are faced with the task of specifying both the complete geometrical structure at the Evaporation Event and the physical laws that obtain there. This is sure to be a non-trivial task, although there are some clues to go on. As noted, it seem [sic] fairly obvious how the light cone structure must be formed at the Evaporation Event... But even with those clues, there may be various plausible ways to specify the physics of the anomalous event (Maudlin, 2017, p. 18).

Manchak and Weatherall (2018) counter that Maudlin has committed an oversight, which stems directly from his over-reliance on Penrose diagrams. Penrose diagrams have several representational limitations. For one, conformally-related metrics have identical Penrose diagrams (in that angles are preserved across the metrics), which is why they obscure the magnitudes of conformal transformations at different points. An additional complexity arises when a pathological metric is conformally related to a well-behaved metric, making it impossible to determine from a Penrose diagram alone whether the spacetime under consideration is pathological or not.

Manchak and Weatherall remark that Penrose diagrams also obscure the trademark feature of black hole evaporation, namely that the event horizon shrinks as Hawking radiation carries away energy. Quite the contrary, Penrose diagrams portray event horizons as expanding at the speed of light. So, to make conclusive inferences about an evaporation spacetime, we have to dispense with Hawking's Penrose diagram, which potentially belies pathological structure, and move to a spacetime diagram.

The one Manchak and Weatherall (2018) present (in Eddington-Finkelstein coordinates) reveals that when the event horizon shrinks to zero radius and a spacetime

[11] I'm acutely aware that Maudlin (2017) disagrees with the conclusion of the general relativistic argument; however, his argument motivating the spacetime structure of Hawking's Penrose diagram is what's relevant for this analysis.

point replaces the singularity, there's a degeneracy of lightcones associated with that spacetime point. Since there is no way of picking one lightcone over the other, the final evaporation event undermines local causal structure and compromises any curve passing through it. Therefore, it also severely and dramatically undermines global causal structure. Thus, they infer that an evaporation spacetime is physically unreasonable.

> The breakdown of causal structure at the "evaporation event" has consequences. For one, it leads to a breakdown of the laws of physics. Consider, for instance, Maxwell's equations... If no consistent metric lightcone can be defined at a point, it follows that Maxwell's equations, as standardly understood, also cannot be defined there (Manchak and Weatherall, 2018, p. 623).

In fact, Manchak and Weatherall (2018) integrate the physical unreasonableness of evaporation spacetimes as a premise into their formulation of the black hole information loss paradox. The contradiction they purport to unmask is that black hole evaporation both is and isn't physically reasonable. They're among the few philosophers who are persuaded by the legitimacy of the black hole information loss puzzle, and they aver that bypassing the contradiction involves rejecting either the formation or complete evaporation of black holes.

This dispute sheds light on methodological fallacies at multiple levels. One level at which this dispute oversteps its reach is in attempting to say something definitive about the causal structure at E. Like Penrose diagrams, spacetime diagrams leave more room for interpretation than either side leads on. Because spacetime diagrams do not preserve angles, lightcones can tilt, change shape, potentially merge, etc. [12] No amount of back and forth will lay to rest the correct interpretation of Penrose and/or spacetime diagrams until they're informed by a family of solutions. Whether or not the final evaporation event is underspecified in its causal structure is open to debate and definitely not "fairly obvious".

The issue of the causal structure at E is separate from ascertaining the status of global hyperbolicity (as defined for a standard relativistic spacetime) in an evaporation setting. Manchak and Weatherall convincingly demonstrate that global hyperbolicity can't be achieved in any rigorous sense regardless of imposing a Lorentzian metric on E. Figure 2.1 reveals how a null ray propagated backwards from $\mathcal{I}^+$ that intersects the final evaporation event doesn't follow a unique trajectory (not depicted); it can be extended along the event horizon or reflected immediately to $\mathcal{I}^-$. This indeterminism holds regardless of the creativity with which this evaporation Penrose diagram is foliated.

Circling back to the original question now, does the tenuously-named general relativistic argument lead to a prima facie acceptable conclusion or not? Despite the stalemate over the predictions of general relativity, I've gathered that the answer is

[12] I'd like to thank Tim Maudlin for explaining his interpretation of the spacetime diagram in private communication.

indeed yes, but not because Maudlin's position won out. Even if E fails to admit of a unique lightcone or any metric whatsoever, Manchak and Weatherall (2018) do not dispute the foliability of an evaporation spacetime into global simultaneity slices. No more than that is needed for the argument to go through because the relevant inference pertains to comparing sets of degrees of freedom contained on early versus late-time global spacelike hypersurfaces, and labeling such surfaces as Cauchy or non-Cauchy has no bearing on the conclusion.

Nevertheless, a deeper level at which this dispute makes a category mistake is its foray into inapplicable domains.[13] Taking Hawking's Penrose diagram at face value, a product of semi-classical gravity, is already implicitly trespassing into the territory of high-energy quantum gravity. Neither general relativity nor quantum field theory can handle the ineffably microscopic, Planck-scale physics in the neighborhood of the final evaporation event or singularity. My stance is that taking Hawking's Penrose diagram at face value is temporarily justifiable to push the framework as far as it can go, which is my aim in engaging charitably with the black hole information loss puzzle.

What's methodologically questionable, however, is making substantive claims about the physical reasonableness of the final evaporation event à la Maudlin without acknowledging quantum gravitational considerations. On the other end of the spectrum, I'm unsure of why the failure of Maxwell's equations at E would prompt Manchak and Weatherall to infer physical unreasonableness and conclude that the laws of physics break down everywhere when classical electromagnetism was never supposed to hold in Planckian regimes.

Consequently, I'm far from persuaded that Manchak and Weatherall (2018) have compellingly cast the black hole information loss puzzle as paradoxical. Their formulation is designed to reject black hole evaporation based on non-global hyperbolicity, which restrains the scope of physical reasonableness. Not only is general relativity ill-suited to determine the physical reasonableness of the final evaporation event, the culprit of non-global hyperbolicity, but physical reasonableness within semi-classical gravity should also be determined in conjunction with quantum field theory.

Therefore, Manchak and Weatherall's conclusion that black hole evaporation is physically unreasonable doesn't pull its weight as prima facie acceptable, which dissolves the contradiction that they try to prop up. Furthermore, their formulation of the black hole information loss paradox mandates a resolution by denying the formation or complete evaporation of black holes. This framing has the methodological vulnerability of rendering the most influential proposals in the discourse, those that maintain complete black hole evaporation, potentially incapable of solving it.

As I anticipated in Section 2.2, rather than forcing a verdict when the jury's still out by insisting on pathological spacetime structure and dynamical incoherence upfront, which is akin to prematurely giving up on the states-plus-laws toolkit, it's more helpful for progress in quantum gravity to assume the validity of Hawking's framework at the outset and subsequently expose an internal contradiction in a reductio ad absurdum

[13] I'd like to thank my advisor, Christian Wüthrich, for bringing to my attention the relevance of high-energy quantum gravity.

argument. By affording the conclusion of the general relativistic argument prima facie acceptability, I take up the challenge to convince sympathizers of this dynamical narrative of another insidious threat lurking underneath.

Given that the general relativistic argument minimizes contributions from quantum field theory, we have to narrow in on the quantum goings-on of matter fields to either corroborate or invalidate the inference that Σ_3's degrees of freedom constitute the global post-evaporation system, which is, of course, the task of the subsequent section. Spoiler alert: Quantum theory is consistent with invalidating the conclusion of the general relativistic argument, allowing me to put forth a more robust formulation of the black hole information loss paradox.

2.4.2 The Quantum Theoretic Argument

The general relativistic argument on its own is too weak to pose a paradox since the dubious reverberations of non-global hyperbolicity fall under the purview of quantum gravity. Evaporating black holes, nevertheless, are not just the product of relativity theory. They're the product of semi-classical gravity, the union of quantum field theory on a back-reacting spacetime. I demonstrate that even if an evaporation spacetime is causally and dynamically well-behaved, the conclusion of the general relativistic argument contradicts that of another, distinct dynamical argument coming from quantum field theory. This contradiction is truly exciting because by uncovering how general relativity and quantum field theory conflict during black hole evaporation, we can make progress in quantum gravity.

In the general relativistic argument, we explored the possibility that the state of late-time radiation is mixed because it's residing in a lower-dimensional Hilbert space. However, we should also investigate the possibility that the final mixture is improper in the original Hilbert space, insinuating that late-time Hawking radiation is an entangled subsystem. Given the physical underdetermination of mixed states, how do we go about figuring this out? We prepare an entangled subsystem at earlier times, evolve it unitarily, and fathom whether we wind up with Hawking radiation. Unitarity is of great help because it keeps constant the extent of external entanglement by conserving von Neumann entropy – as long as the subsystem is embedded in a closed system, such as the entire universe. As such, self-contained systems always stay isolated from their environment, and entangled subsystems always stay correlated with their environment.

If we're handed the initial data of a pure state with internal entanglement, then we can take advantage of unitarity through the following procedure. The pure state reveals the total referent system, i.e., complete set of degrees of freedom and nonlocal correlation structure. However, we can choose to neglect or trace out a subsystem in favor of analyzing the remaining, referent subsystem, comprised of a proper subset of degrees of freedom. By stipulation, our choice of a partial trace leaves us with an externally entangled subsystem described by a mixed state. There's no ambiguity in the interpretation of that mixed state – we prepared it that way (see Susskind 2012;

Polchinski 2017).

Now, by acting on that mixed state with a unitary time-evolution operator and refraining from concurrently acting on the pure state of the whole system, the output state can be nothing but mixed and entangled. The presence of entanglement necessitates that the output state is still describing a subsystem whose degrees of freedom are entangled with those of the traced out subsystem. We are – by design – ignorant of this subsystem's nonlocal correlations with the traced out subsystem; for that information, we'd need to evolve the overarching pure state unitarily. But because we've already fixed the referent system, unitary evolution leaves no room for doubt that generating a succession of mixed states continuously describes the entanglement of that subsystem.

Let's contemplate Figure 2.3 and see what happens when we perform the aforementioned procedure on Σ_1. Σ_1 is a pure vacuum state of two Hawking pairs. The purple entanglement curves implicitly represent the exact form of the nonlocal correlation structure between the positive and negative-energy degrees of freedom. After tracing out the negative-energy members, i.e., blue nodes, we're left with the positive-energy members, i.e., red nodes, surrounded by red dotted ovals. Call this segment Σ_{1+}, which describes the referent entangled subsystem.

Σ_{1+} is a mixed state omitting information about how the positive-energy quanta are entangled with their negative energy partners. It's necessarily derived from Σ_1 because it relays partial information about a proper subset of degrees of freedom. But only Σ_1, a global pure state, encodes the complete information. Furthermore, we're aware from Section 2.3.2 that Hawking pairs are monogamously entangled. Therefore, Σ_{1+}'s von Neumann entropy is maximized via uniform probability distributions, which correlate the occupation numbers of positive and negative-energy quanta to maintain Σ_1 as a zero-energy vacuum state.

Recall from the previous section that evolving Σ_1 unitarily for a preset period of time leads to Σ_2. Similarly, evolving Σ_{1+} unitarily by that same period of time leads to Σ_{2+}, the segment of Σ_2 with only positive-energy, red nodes. As is evident in the diagram, the red nodes are still connected by purple curves to the blue nodes on Σ_2, but the blue nodes are not included in the evolution. The referent entangled subsystem is fixed – the Hawking radiation we know and love.

That's the dynamical backstory behind applying the Bogoliubov transformation on Σ_{2+} in order to solve for the post-evaporation coefficients of creation and annihilation operators (see Hawking 1975; Hawking 1976; and also Wallace 2018 for an exposition). Subsequently, Fredenhagen and Haag (1990) produced an explicitly time-dependent calculation relating short-distance, near-horizon behaviors to so-called observation regions both moderately and extremely distant from the black hole. Their result not only corroborates Hawking's predicted thermal spectrum (discounting deviations arising solely from gray-body factors), but it also attests to the tame dynamical evolution from Σ_{1+} to Σ_{2+}.

That said, notice how Σ_{2+} bears a striking resemblance to Σ_3. It contains the same set of positive-energy degrees of freedom in a mixed state. My aim is to chari-

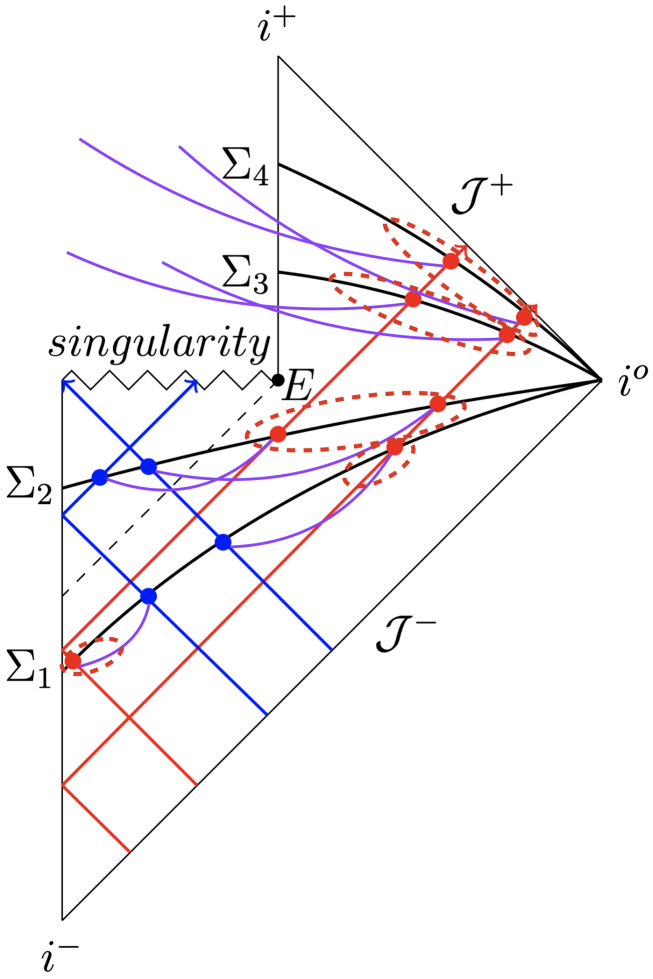

Figure 2.3: Preservation of Exterior Entanglement
Σ_{1+} and Σ_{2+} are entangled subsystems of positive-energy degrees of freedom, i.e., red nodes surrounded by red dotted ovals. They evolve unitarily into Σ_3 and Σ_4, whereby entanglement is preserved.

tably engage with the black hole information loss puzzle, so whenever conditions that are conducive to unitarity obtain, like the conservation of degrees of freedom, the preservation of state properties, and well-behaved local spacetime structure, then I'm presupposing it's the default, which is to say that we should strive to evolve from Σ_{2+} to Σ_3 unitarily. After all, Hawking conceptualized black hole radiation as energy flux across the event horizon eventually felt at future null infinity $\mathcal{I}^+$.

But how should we handle the non-unitary hiccup at the final evaporation event, E? For now, it doesn't concern us. Even though the transition from Σ_2 to Σ_3 is pure-to-mixed due to interference from the black hole singularity, there's no obvious reason we can't smoothly evolve Σ_{2+} past E, which encounters no such interference. Belot et al. (1999) affirm that "this [global] failure of unitarity need not imply that the local laws of field propagation have been altered" (p. 200). They insist that in sufficiently small neighborhoods, conditions conducive to unitarity obtain.

In addition to the prospect of quantum fields evolving unitarily outside the black

hole, when we further take into account that unitarity conserves von Neumann entropy, we realize that Σ_3 must describe late-time Hawking radiation as an entangled post-evaporation subsystem. Moreover, since we've established in Section 2.4.1 that Σ_3 evolves unitarily into Σ_4, it must be the case that any mixed post-evaporation state necessarily describes late-time Hawking radiation as an entangled subsystem.

We can now satisfactory explain the presence of the purple entanglement curves in Figure 1.17. Though the parallels between Σ_{2+} and Σ_3 as entangled subsystems were already, on the face of it, highly suggestive, informing the vast majority of live proposals as I summarize in Section 2.6, I've credibly backed up this hunch in an original contribution to the literature through the quantum theoretic argument.[14]

The quantum theoretic argument:

1. The degrees of freedom of Σ_{2+}, a mixed pre-evaporation state, constitute an entangled subsystem.

2. Σ_{2+} evolves unitarily to Σ_3, a mixed post-evaporation state.

3. Unitary evolution preserves entanglement from Σ_{2+} to Σ_3.

4. Only subsystems are entangled.

Conclusion:

5. Therefore, the degrees of freedom of Σ_3 constitute an entangled post-evaporation subsystem.

Taken in isolation, nothing about this conclusion raises alarms. The laws of standard, unitary quantum theory straightforwardly add their own dynamical layer to the feature film of black hole evaporation. Unlike in the general relativistic argument, where we ascertained the loss of something important – degrees of freedom, here we've counterintuitively ascertained the conservation of something important – external entanglement. Pure-to-pure and mixed-to-mixed transitions (involving density matrices of the same collection of vectors) provide a pretty hefty ontological guarantee. Once a referent system is fixed, its fundamental composition, possession of core properties, and relationship to other systems are all encoded and immortalized over the course of unitary evolution (presuming no major shocks to the system at large). It's tantalizing to proclaim that black hole evaporation knows how to keep track of and protect the subsystem to which it owes its namesake – Hawking radiation.

Speaking of which, let me respond to potential objections targeting the third and fourth premises doing the heavy-lifting. I recognize that they're susceptible to denial because they're not ironclad. There's no proof that I or anyone can give establishing the entanglement of mixed radiation states without deriving them as reduced density matrices from pure global states. Yet as has been emphasized repeatedly, the issue

[14] For an extended version of this argument, see Appendix B.

is that there are no pure global states to be had after the final evaporation event. This situation does expose the fragility of the quantum theoretic argument, and that fragility is an acceptable and unsurprising revelation, as the impending conflict with the general relativistic argument ushers the downfall of either or both arguments. To reiterate the sentiment I led early on, the contradiction that I'm foreshadowing is best utilized as a retroactive window into the landscape of purported solutions.

The fact that purported solutions not only exist but thrive in quantum gravity research entails that there must be weak links, but what I aim to drive home is that challenging the third and fourth premises of the quantum theoretic argument to escape the anticipated paradox is not the lowest-hanging fruit. Though the account of unitary evolution preserving entanglement across Σ_{2+} and Σ_3 isn't definitively true, it isn't definitively false by the same token. The absence of pure global states prevents recourse to a precise, mathematical proof settling *any physical interpretation* of mixed radiation states. For example, casting them as global macrostates also requires accommodating global microstates, but again, the pure global states which are supposed to fulfill this role are incompatible with Hawking's framework up to perturbative corrections, as I delved into in Chapter 1 (see Mathur 2009; Wallace 2020). Hawking (1976) himself added to the underdetermination problem by redefining mixed states as generic microstates and inventing the superscattering formalism to supersede global unitarity in the first concrete application of density matrix realism.

A major component of the information loss puzzle is that the dynamical history of black hole evaporation sheds little light on answering why radiation states remain mixed post-evaporation. In spite of that ambiguity, I'm holding my ground advocating that the third and fourth premises do a better job than their negations. Although a rigorous demonstration of sustained entanglement between Σ_{2+} and Σ_3 is not in the cards, I'll begin planting the seeds for this position's plausibility, which is enough to provisionally accept the conclusion of the quantum theoretic argument and gain valuable insights from a didactic paradox.

To that end, let's walk through and relieve some of the pressure points. The strongest objection I foresee is that unitary evolution certainly preserves mixedness from Σ_{2+} to Σ_3, but it doesn't necessarily preserve entanglement. At first glance, such an objection seems incoherent given that unitary evolution almost always conserves von Neumann entropy. But it's not if we recognize we're equivocating on the meaning of 'entanglement'.

The third premise is consistent with two readings of 'entanglement'. First, entanglement is a mathematical property of vectors from which a specific type of density matrix is derived, that without a sharply peaked probability distribution. According to this definition, entanglement would be equated with mixedness and undeniably preserved under unitary evolution. The fourth premise, however, could be false since the physical interpretation of a mixed state is underdetermined. Such a move is unattractive for quantum state realists who distinguish between the nonlocal correlations of entanglement and the local correlations of ordinary statistical ensembles. Anti-realists, such as Jaynes (1957b), would conversely be more than happy to collapse the two under

a single epistemic banner.

I argued from the beginning that the anticipated paradox of phantom entanglement is a game to be played by quantum state realists. Thus, the second, more conventional reading of entanglement treats it as a physical relation encoding nonlocal correlations (see Calosi and Morganti 2021). That's how I've been treating the term, especially when I've explicitly distinguished between self-contained systems and systems with external entanglement. It's this background assumption that lends support to the inference that positive von Neumann entropy quantifying the extent of external correlations is reserved for entangled subsystems. Seeing as unitary evolution generally conserves von Neumann entropy, it all but ensures that Σ_3 is an entangled subsystem.

However, one may counter that those general conditions may not hold in black hole evaporation. Most notably, the global state is eventually disturbed when the topological discontinuity at the final evaporation event introduces a major shock by inducing a pure-to-mixed transition, leaving room to sever the entanglement – though how much room is precisely the question. Clifton and Halvorson (2001) show that only non-unitary operations on the global state are capable of eliminating entanglement among subsystems (partitioned by a chosen Hilbert-space factorization), the pedestrian example being that of projection.

The thought goes that because the black hole singularity is a global feature of spacetime, and its presence is inextricably linked to global non-unitary evolution, then it must be capable of severing the entanglement between Hawking radiation and the black hole interior despite the radiation state remaining mixed onward. Therefore, by capitalizing on the physical underdetermination of a mixed state, one could reject the third premise establishing the preservation of entanglement relations under unitarity without compromising the preservation of state mixedness under unitarity.

The information measure that usually gets labeled 'von Neumann entropy' would have to be conserved throughout, but that label would eventually cease to be appropriate. What resembles von Neumann entropy must conceptually shift into something else, a new type of entropy, and that information measure would be conserved thereafter. A single mathematical object would thus represent two distinct physical situations along the continuous evolution from Σ_{2+} to Σ_3. Up until E, the density matrix representing Hawking radiation would imply the presence of entanglement relations with a complementary, traced out subsystem, and von Neumann entropy would be conserved. After E, the density matrix would imply the opposite – the absence of entanglement relations with a complementary, traced out subsystem, wherein a different entropy would be conserved.

While refuting the account of sustained entanglement across Σ_{2+} and Σ_3 is certainly an option, it's far from inevitable or the most bankable deterrent against a looming paradox. Even though the singularity is indeed a global feature of spacetime, it's not obvious that it should correspond to a global, non-unitary operation on quantum fields. What's happening is that as quantum fields propagate, the entangled subsystem localized inside the black hole encounters a singular hiccup preventing it from propagating any further, confining the interference to a bounded spatiotemporal

region. The complementary entangled subsystem outside the black hole, conversely, carries on business as usual.

I argued in the previous section that annihilation is better conceived as a spatiotemporal process, in which the disappearance of black hole degrees of freedom can only be articulated by spatiotemporal observers post-elimination, i.e., observers in the exterior region. In that vein, a case could be made that the change to the global state arises not by intervening on it wholly but by intervening on it partially to eliminate the entangled black hole subsystem. As Clifton and Halvorson (2001) prove, no local operation, unitary or not, can neutralize entanglement in quantum field theory.

This analysis tracks diverging attitudes towards the question, where is the spacetime singularity? The straightforward answer is that it's nowhere because by definition, singularities aren't part of the manifold. Nevertheless, it would be perfectly sensible to answer that it's at the center of a black hole, since only by crossing the event horizon and approaching $r = 0$ could something confront the singularity. Because both answers are correct, it's no wonder that quantum field theory can't decide whether to package the singularity's annihilating effects as a global or local operation, and by extension, whether to sever the entanglement at the final evaporation event.

Thus far, neither the veracity nor the falsity of the third premise has the upper hand, and frankly, the stalemate is not important. Just the possibility of the third premise being true justifies at least entertaining the quantum theoretic argument to see what the paradox of phantom entanglement can teach us when reverse-engineering the many avenues to forestall it. Yet for the more ambitious aim of tipping the scales for or against it, we'd have to rely on extra-theoretical factors.

Advocating in favor of the third premise, let me briefly lay out several dialectical and philosophical advantages to having mixed radiation states encode residual entanglement. There's a methodological advantage to explain a crucial consideration adopted by the vast majority of contemporary proposals endeavoring to resolve information loss, which I will go into in Section 2.5 and Section 2.6 while taking an even deeper dive in Chapter 3. Despite these proposals being born to naively restore unitarity for unitarity's sake, they essentially take a stand on how to handle the excess entanglement post-evaporation, a strategy that's intricately connected to taking a stand on the interpretation of Bekenstein-Hawking entropy in quantum gravity. Recognizing the third premise as deserving of independent motivation to accomplish what the current discourse already cares about but through a more rigorous route engenders significant gains. That independent motivation is backed up by a suggestive philosophical advantage regarding representation relations between the mathematical formalism and reality.

As I emphasized in Chapter 1, there's long been underappreciated precedent for casting black hole information loss as the increase in global von Neumann entropy (see Page 1995; Mann 2015), which hangs on the conservation of the exterior subsystem's von Neumann entropy. Construing von Neumann entropy as entanglement entropy throughout simplifies the representation relation between density matrices and physical systems to describe black hole evaporation. Since Hawking pair production

already invokes mixed states to represent positive-energy partners entangled with their negative-energy counterparts, maintaining that representation as those mixted states unitarily evolve into a system recognized as Hawking radiation pins down their physical interpretation when no funny business is taking place dynamically. In standard quantum theory, unitary evolution is the epitome of no dynamical funny business.

I side with Maudlin (2018) in appealing to one-to-one representation relations between mathematical and physical entities, with the caveat of expressing that preference whenever they're available and the most explanatorily useful. For anyone who thought the twofold representation relation between mixed states and physical systems leaves much to be desired, at least unitarity hasn't thus far messed with it in the middle of the operation when the subsystem at hand is left alone.

However, if entanglement genuinely snaps due to the singularity, nothing about the smooth, unitary evolution of the mixed state would reveal when the switch is flipped. In some contexts, we could zoom out and trace that 'physical snap' to a global projection operation which legitimately acts on all subsystems. However, Hawking radiation is unsurprisingly left alone by the singularity, so the mechanism behind its reaction to the 'physical snap' is much more difficult to accommodate. Ergo, from a philosophy of science perspective assessing theoretical virtues, I'm hesitant that denying the third premise is conducive to a fruitful framework for black hole evaporation.[15]

Any reluctance aside, I've taken great pains to flesh out intuitions that unitary evolution preserves entanglement from Σ_{2+} to Σ_3. For the pedagogical purpose of presenting a paradox and then subsequently defusing it, an exercise which cuts through the noise in the debate and taxonomizes all information-restoring proposals based on a powerful guiding principle bearing on black hole thermodynamics and statistical mechanics, I hereby part ways with honorary density matrix realist Stephen Hawking (who incidentally changed his mind later in life in Hawking et al. 2016) and grant the position that the quantum theoretic argument leads to a prima facie acceptable conclusion.

2.5 A Kinematic Clash: The Paradox of Phantom Entanglement

The astute reader will have foreseen my punchline coming from a mile away. The conclusions of the general relativistic and quantum theoretic arguments subvert each other, making the black hole information loss puzzle credibly paradoxical. This problem is kinematic in nature as the tension revolves around a decisive state in black hole evaporation, Σ_3, thus setting the tone for subsequent states, like Σ_4. What's at stake is entanglement between physical and seemingly unphysical, i.e., 'phantom' degrees of freedom, which is why I've rebranded the black hole information loss paradox, the

[15]Much of my defense of the quantum theoretic argument is inspired by conversations with Nick Huggett and Sam Fletcher, to whom I'm very grateful for prodding me to sharpen my views.

"paradox of phantom entanglement".

Paradox of Phantom Entanglement:

1. The degrees of freedom of Σ_3 constitute a global post-evaporation system.
2. The degrees of freedom of Σ_3 constitute an entangled post-evaporation subsystem.
3. The degrees of freedom of Σ_3 constitute Hawking radiation.

Conclusion:

4. Therefore, Hawking radiation is a global post-evaporation system and an entangled post-evaporation subsystem.

The terms 'global system' and 'subsystem' are diametrically opposed and mutually exclusive. The universe only has room for one global system, whereas partitioning the global system entails at least two subsystems, so a global system cannot simultaneously be a subsystem. Therefore, we've encountered a blatant contradiction between general relativity, which attests to late-time Hawking radiation the sole survivors of black hole evaporation, and quantum field theory, which attests to its external entanglement. Given that both conclusions are prima facie acceptable and pull their weight in the contradiction, the paradox of phantom entanglement is worthy of its title and highlights precisely where the internal disagreement lies within the semi-classical framework.

To finish this critical assessment and further develop the phantom entanglement metaphor to that end, I wish to bring the two arguments into contact and pinpoint where they clash. As we know by now, the quantum theoretic argument entails that late-time Hawking radiation is an entangled subsystem. And in Chapter 1, I commented that the mixed states of entangled subsystems are parasitic on pure states. This parasitic relationship is analogous to ontological dependence for abstract entities – in order for the mixed state in question (with positive von Neumann entropy) to exist, there must exist a pure state (with zero von Neumann entropy) from which it's derived. Translated physically, in order for the entangled subsystem of interest to exist, there must exist a global system supplying the nonlocal correlations (see Calosi and Morganti 2021).

Consequently, there must also exist a complementary entangled subsystem that's been traced out, whose degrees of freedom complete the global system. If we're to take the quantum theoretic argument at face value, we should be able to locate the complementary entangled subsystem. How do we do that vis-à-vis Figure 2.4? The first step is to identify spacelike separated degrees of freedom that are simultaneous with the original subsystem in the relevant foliation.The second step is to ensure that there are sufficiently many entangled degrees of freedom in the complementary subsystem to 'purify' the original one. In other words, both subsystems must be equally entangled

with each other with identical Boltzmann and von Neumann entropies, wherein their storage capacities are sufficient to hide information in nonlocal correlations (see Page 1993a; Susskind 2012).

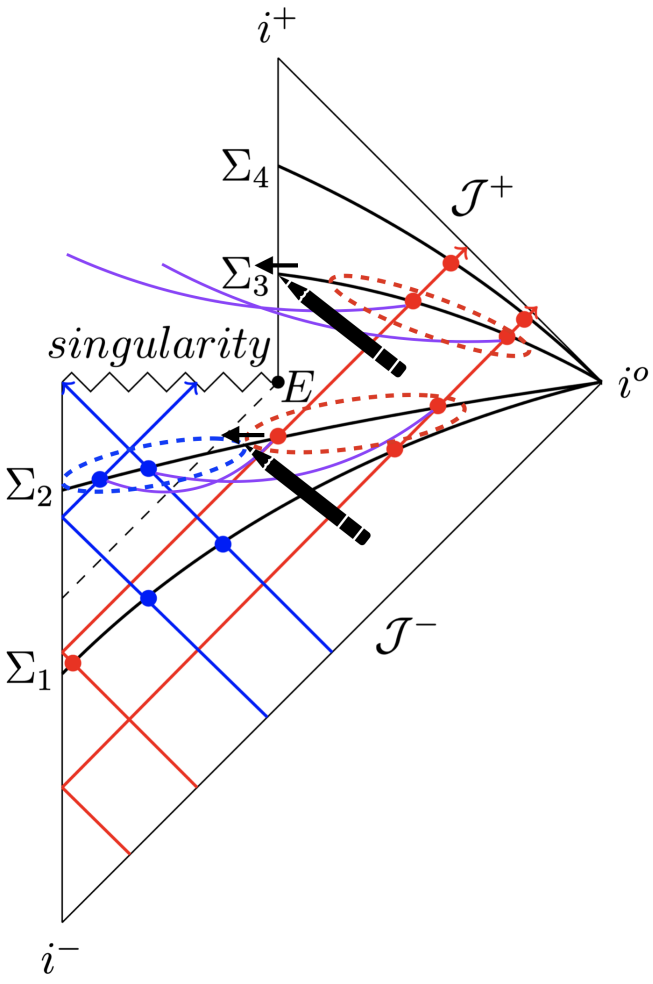

Figure 2.4: Phantom Entanglement

Σ_{2-} and Σ_{2+} are partial Cauchy surfaces with (negative-) positive-energy degrees of freedom, i.e., (blue) red nodes surrounded by (blue) red dotted ovals. They're both complementary entangled subsystems with equal Boltzmann and von Neumann entropies. Σ_3, on the other hand, has no complementary entangled subsystem because the surface cannot be extended.

Before we execute these steps for Σ_3, it would be beneficial to do so for Σ_{2+}. Recall that Σ_{2+} contains just the entangled positive-energy degrees of freedom of Σ_2. On the diagram, it's the right-half segment of Σ_2 from the event horizon practically to spacelike infinity, i°. We see a red dotted oval surrounding two red nodes, each connected to a purple curve with a thus far empty node slot on the other end, signifying entanglement with mysterious degrees of freedom. The fastest strategy to execute both steps is to imagine placing a pen on the left edge of Σ_{2+} (the event horizon) and tracing it along the rest of Σ_2 (for the sake of simultaneity) as if we're continuously extending Σ_{2+} into the black hole.

First, we hit a blue node. Excellent – one of the purple curves has its empty node slot occupied, which implies we've found an entangled negative-energy partner and reinstated a Hawking pair. But we haven't yet identified sufficiently many entangled degrees of freedom. The second red node is connected to a purple curve that still has an empty node slot on the other end. In other words, both subsystems' Boltzmann and von Neumann entropies don't match. So, we go on tracing until we hit the other blue node, which basically brings us to $r = 0$. Now both purple curves have occupied node slots, and we've reinstated the second Hawking pair.

After having performed this procedure, we learn that the complementary entangled subsystem must be Σ_{2-}, the left-half segment of Σ_2 behind the event horizon in a bounded spatial region. Σ_{2-} is comprised of negative-energy degrees of freedom, two blue nodes surrounded by a blue dotted oval, located inside the black hole. The Boltzmann and von Neumann entropies of Σ_{2-} certainly match that of Σ_{2+}. Although the union of Σ_{2+} and Σ_{2-} recovers Σ_2 as a geometric Cauchy surface, it doesn't restore the nonlocal correlations within the global system. Therefore, Σ_2 as a pure state has implicitly been in the background this entire procedure.

Let's move on to Σ_3. Like Σ_{2+}, it contains just entangled positive-energy degrees of freedom. Again, we see a red dotted oval surrounding two red nodes, each connected to a purple curve with an empty node slot on the other end, signifying entanglement with mysterious degrees of freedom. Imagine placing the pen on Σ_3's left edge at $r = 0$ to continuously extend it. But wait – we can't continuously extend it. We're stuck. We've hit the edge of the diagram itself. This is the juncture at which the general relativistic argument, especially non-global hyperbolicity, comes into play.

Both Σ_2 and Σ_3 span the maximal radial distance: from $r = 0$ to i°, even though the former is a Cauchy surface and the latter is not. Unlike for Σ_{2+}, there's no more space or degrees of freedom available to host a complementary entangled subsystem for Σ_3. Everything in the ontology at that moment of time has been exhausted. What would've been the complementary entangled subsystem – the black hole interior – has disappeared post-evaporation. Hence, the purple curves on Σ_3 representing entanglement relations can never have all of their node slots occupied. Late-time Hawking radiation seems to be entangled with degrees of freedom that are nonexistent in spacetime, like phantoms.

To summarize, the quantum theoretic argument insists that we embed Σ_3 into a genuine Cauchy surface associated with a pure state. Yet the general relativistic argument asserts that we can't because the singularity has annihilated the complementary entangled system. The contradiction is an unheard of kinematic problem of entanglement obtaining without state construction sanctioned by the underlying theory! After the disappearance of the event horizon, global states cannot even exist, in some sense. It's therefore no longer possible to ascribe any physical state to the paradoxically global system of entangled Hawking radiation. The concurrent elimination of degrees of freedom and appearance of external entanglement for a global system are not just *adjacent* to being physically and metaphysically unintelligible – as I asserted in Chapter 1 – they *are* physically and metaphysically unintelligible.

One might object that the paradox of phantom entanglement is a foliation-dependent pathology that only comes about through a bad choice of simultaneity slices, in which the standard scheme fails to consider legitimate spacelike hypersurfaces containing both the black hole interior and late-time Hawking radiation.[16] In a similar vein, one might protest that the paradox of phantom entanglement dissolves when we remove the crutches of global simultaneity and persistence claims for its formulation. The legitimacy of the paradox suffers if it's not formulated in the language of invariants; otherwise, it risks importing controversial metaphysical theses about time, such as presentism or endurantism.[17]

I have two responses to these objections. As a first pass, I'm cautiously optimistic that the technicalities could be worked out to render the phenomenon of phantom entanglement more stubborn than what these objections let on. The rules of the game can be updated to concretely deliver phantom entanglement by prodding Hawking radiation to the future lightcone of the final evaporation event E (refer to Figure 2.2), such as by trapping massless Hawking radiation and transporting it along timelike trajectories or by considering massive Hawking radiation.[18]

Should the above scenarios pan out, dispensing with simultaneity slices would need to be reworked in the framework of algebraic quantum field theory. Then, the contradictory conclusion would be recoverable by comparing the algebra of observables on pre- and post-evaporation spacetime regions without recourse to temporal parametrization. Therefore, the gravitas of the kinematic incompatibility arising out of a clash between the general relativistic and quantum field theoretic arguments could transcend the arguments' cursory reliance on temporal considerations.

But even if one doesn't accept the terms of the game I'm playing in order to produce a paradox, and even if it turns out that the mainstream narrative (catalyzed by Hawking himself) has exploited crying wolf by decrying global non-unitarity over an unnecessary foliation scheme or evaporation diagram, thus potentially obsolescing phantom entanglement along with influential reactions and developments within quantum gravity, preemptive strikes against a semi-classical paradox (that I've transparently advertised as a temporary) doesn't absolve us from the responsibility to answer the question it brings to fore. Is late-time Hawking radiation in a mixed state as a genuinely entangled subsystem or as a disentangled global system?

Either answer commits us to a speculative school of thought about black hole thermodynamics and statistical mechanics, each fraught with controversy, as I'll unveil in Chapter 3. The impending philosophical minefield thus undermines the utility of loopholes to the paradox arrived at primarily on technical grounds. You may agree with the loopholes yet disagree with the unintended consequences, especially those attenuating the celebrated proportionality relation between horizon area and black

[16] I'd like to thank Nick Huggett and Dominic Ryder for engaging with me on this technicality.

[17] I'd like to thank Chris Wüthrich, Baptiste le Bihan, Nick Huggett, Hans Halvorson, and David Wallace for invaluable debate on separate occasions regarding the dispensability of global simultaneity and the role of metaphysical theses about time.

[18] My heartfelt thanks goes out to Dominic Ryder for suggesting these ways out.

hole entropy. It's widely undisputed that a satisfactory determination of the state describing Hawking radiation prompts modifications to Hawking's evaporation diagram anyway to incorporate Planck-scale corrections, and my larger purpose is to ultimately draw insightful, big picture connections between the original conceptualization of information loss and ensuing paradigm shifts centered around the meaning of black hole entropy in quantum gravity.

2.6 Reifying Phantoms in Quantum Gravity

As promised, I've delivered a plausible reductio ad absurdum argument culminating in a black hole information loss paradox, what I've coined the paradox of phantom entanglement. An immediate reaction might be to throw your hands up in the air and forget about Hawking radiation, black hole evaporation, and semi-classical gravity, a demonstrably self-destructing framework. While this reaction would be warranted logically, it would also be too quick. Peering ahead towards quantum gravity, Hawking (1975) expresses incisive physical intuitions that the quantization of the metric naturally insinuates radiating and evaporating black holes.

> It should not be thought unreasonable that a black hole, which is an excited state of the gravitational field, should decay quantum mechanically and that, because of quantum fluctuation [sic] of the metric, energy should be able to tunnel out of the potential well of a black hole (Hawking, 1975, p. 202).

With quantum gravity paving the way for black hole evaporation, as well as the assumption that Hawking's calculation becomes unreliable before the final evaporation event when the invocation of novel physics is past overdue (see Crowther et al. 2021), it's worthwhile to bring black hole evaporation into the jurisdiction of quantum gravity. Then multiple approaches, not least the dominant paradigms of string theory and loop quantum gravity, should spearhead the initiative to re-derive Hawking radiation with black hole thermodynamics and statistical mechanics as the end goal (see Wallace 2019). Judging by the surfeit of proposals, intellectual investment in a resolution isn't subsiding any time soon.

A successful re-derivation of black hole evaporation foregoes ones of the contradictory conclusions, thereby reifying phantom degrees of freedom and reinserting them into spacetime. On the one side, if Hawking radiation remains an entangled post-evaporation subsystem, then a black hole remnant must persist to accommodate the complementary entangled subsystem (see Hossenfelder and Smolin 2010; Chen 2020). On the flip side, if only Hawking radiation survives as the global, post-evaporation system, then it must be self-contained, likely with internal entanglement relations (see Page 1993b; Susskind et al. 1993; Mathur 2009). A commitment to one or the other conclusion is obligatory for physical coherence even if quantum gravity replaces the general relativistic and/or quantum theoretic argument that originally led us there.

2.7 Conclusion: Black Hole Information Loss is Paradoxical

To conclude, I want to offer final reassurance that I've faithfully reproduced the black hole information loss puzzle culminating in the paradox of phantom entanglement. Pay attention to the following passage by Belot et al. (1999), which artfully and concisely weaves together the general relativistic and quantum theoretic arguments:

> The density matrix ρ_{ext} associated with the region exterior to a black hole at a time Σ_2 will describe a mixed state ($(\rho_{ext})^2 \neq \rho_{ext}$). This is because ρ_{ext} is obtained by tracing out over degrees of freedom describing the interior of the black hole, and because the exterior and interior degrees of freedom are correlated – in particular, Hawking shows that the radiation propagating out towards spacelike infinity is correlated with the radiation entering the black hole. Of course, the mixed character of ρ_{ext} at Σ_2 is unexceptional. For ρ_{ext}, describing a proper subsystem of the total system, is compatible with the total state which remains pure. But consider what happens after the black hole has evaporated. Now the state of ρ_{ext} is just the state of the entire universe. Mixed until the time of complete evaporation ρ_{ext} remains mixed thereafter. So the state of the universe, originally pure (or so we assume), is now mixed (pp. 194-5).

Belot et al. (1999) patently hadn't sniffed out any inconsistencies based on what they say next.

> [T]he pure-to-mixed transition hardly seems to merit what can only be described as the measures of desperation some would adopt to avoid it (p. 221).

Unruh and Wald (2017), physicists and vocal skeptics of the black hole information loss paradox, corroborate that sentiment of incredulity.[19]

> [The] loss of information in black hole formation and evaporation does not violate any fundamental principles of physics and is not, in any way, a radical proposal (p. 16).

It's abundantly clear that the contradiction I've exposed between these two statements: "[A]fter the black hole has evaporated... the state of ρ_{ext} is just the state of the entire universe," and "Mixed until the time of complete evaporation, ρ_{ext} remains mixed thereafter", has slipped under the radar for decades. So, to respond to the repeated allegation that sociological forces have counterproductively distorted the black hole information loss discourse, I fervently agree – not because group think has roused unwarranted puzzlement, but instead, because it has unduly suppressed "measures

[19] I've since learned in private communication that Bob Wald conceives of information loss epistemically.

of desperation" when they were called for the most, to critically assess the unphysical implications of a global radiation state whose mixed origins stem from external entanglement.

As it turns out, the impact of this critical assessment goes beyond the exposure of phantom entanglement. The framing and resolution of the paradox props up a question of paramount importance in quantum gravity: What is the information storage capacity of a black hole? If Hawking radiation is an entangled post-evaporation subsystem, then the black hole's information storage capacity must at least be as large as the radiation's von Neumann entropy for the remnant to serve as a safehouse for the complementary entangled subsystem. Conversely, if Hawking radiation is a global post-evaporation system, then the black hole's information storage capacity must be limited because the radiation's von Neumann entropy vanishes, thereby forcing the evacuation of trapped degrees of freedom.

Opinions about the answer to this question are inextricably intertwined with the interpretation of Bekenstein-Hawking entropy and also the driving force behind the proliferating proposal space. I'm going to argue in Chapter 3 that the dichotomy of safehouse versus evacuation solutions maps onto two umbrella interpretations of Bekenstein-Hawking entropy, causal entropy versus holographic entropy, revealing profound metaphysical ambiguity in the very identity of a black hole, with complications for thermodynamic and statistical mechanical behaviors.

Chapter 3

Facing the Phantom Music: Black Hole Entropy Guides Information Conservation

3.1 Introduction: Spoiled for Choice

Planckian remnants. Baby universes. White holes. Catastrophic membranes. Soft hairs. Wormholes. Holograms. Fuzzballs. The list of prospective solutions to the black hole information loss paradox goes on and on.

Despite innovations in theoretical black hole physics, we're not much closer today than we were 30 years ago to narrowing down a solution. New ideas ineluctably crop up, and then they're added to extensive literature reviews, underscoring how spoiled we are for choice (see Preskill 1992; Page 1995; Belot et al. 1999; Wald 2001; Mathur 2009; Hossenfelder and Smolin 2010; Carlip 2014; Mann 2015; Unruh and Wald 2017; Polchinski 2017). While these literature reviews are excellent, consolidated sources to learn about many proposals in one place, their problem is that they don't organize the proposals to help compare and assess them at a higher level of abstraction. I've certainly gotten lost in the weeds trying to remember the ins and outs and the pros and cons of individual proposals.

In this chapter, I aim to do one better for the discourse. I group the leading proposals into the following two mutally exclusive categories. By the time an evaporating black hole reaches Planck mass, either 1) Hawking radiation is maximally entangled with the interior, or 2) Hawking radiation is barely entangled with the interior. I label proposals belonging to the first category 'safehouse solutions' because ostensibly missing information has been safely stored inside the black hole. I label proposals belonging to the second category 'evacuation solutions' because hidden information has been instead evacuated to the exterior. This dichotomy follows organically from the paradox of phantom entanglement, which I set up and fleshed out in the previous chapter.

Admittedly, I'm now guilty of writing yet another literature review compiling var-

ious options on the table, but I do so to carefully classify them in the aforementioned taxonomy. Once the groundwork is laid, it becomes much more tractable to juxtapose safehouse and evacuation solutions for the purpose of expedient evaluation. I demonstrate how the most powerful feature to compare between safehouse and evacuation solutions puts to rest Wallace's discontent with the mainstream narrative of black hole information loss. He calls it the "evaporation-time paradox" and shrugs it off for belaboring the cost of non-unitary evolution from pre-to-post-evaporation. In order for the black hole information loss paradox to earn its title, he affirms, it has to impact black hole statistical mechanics.

> A much more compelling paradox arises when Hawking radiation is considered not just in the light of quantum mechanics in general, but in particular in the light of black hole statistical mechanics (Wallace, 2020, p. 220).

Although the paradox of phantom entanglement is a cleaned-up version of the evaporation-time paradox, I've devised it precisely to put black hole statistical mechanics in the spotlight. I claim that the disagreement between safehouse and evacuation solutions boils down to the statistical interpretation of Bekenstein-Hawking entropy, i.e., thermodynamic black hole entropy proportional to horizon area. Safehouse solutions attribute Bekenstein-Hawking entropy solely to horizon degrees of freedom. Their primary justification is what I frame as the Causality Argument (CA). Since the event horizon is a causal barrier, only surface degrees of freedom influencing outside regions of spacetime can account for external thermodynamic interactions. Yet total black hole entropy, including causally inaccessible degrees of freedom, is potentially unbounded. Evacuation solutions, in contrast, attribute Bekenstein-Hawking entropy to all black hole degrees of freedom. Their primary justification is what I summarize as the Holographic Principle (HP). Black holes are posited to be finite, ergodic, quantum statistical systems that behave according to the minimal model of the Page curve. Total black hole entropy is thus bounded above by Bekenstein-Hawking entropy. Since it scales with area and not volume, Bekenstein-Hawking entropy represents a holographic bound on entropy density with non-localizable degrees of freedom.

Framed this way, black hole information loss had borne the undercurrents of thermodynamics and statistical mechanics for all major proposals even before the more "compelling" paradox that Wallace alludes to, the Page-time paradox, had taken center stage for a select few (i.e., evacuation solutions). The paradox of phantom entanglement integrates in its resolution the three most urgent questions about black hole physics in semi-classical and quantum gravity, exactly as the founder of black hole entropy himself – Bekenstein – envisioned.

> Three intricately related issues have characterized black hole thermodynamics for the better part of two decades: the meaning of black hole entropy, the mechanism behind the operation of the generalized second law, and the information loss puzzle... The three issues are actually one in the sense that when people find out how to fundamentally resolve one of them, they will have resolved all three (Bekenstein, 1994, pp. 1-2).

The punchline of this (dense) chapter is to construct the scaffolding that orchestrates how "the meaning of black hole entropy, the mechanism behind the generalized second law, and the information loss" play off each other. In Section 3.2, I argue that only a commitment to the generalized second law and its reduction to quantum gravitational degrees of freedom justifies resolving phantom entanglement, after which I explore how modifying Hawking's original framework leads to the classification of safehouse and evacuation solutions. In Section 3.3, I scan the most influential safehouse solutions (pretty much black hole remnants in all shapes and sizes), summarize the reception in the discourse, and analyze their causal implications for Bekenstein-Hawking entropy. I then run the same procedure for evacuation solutions in Section 3.4. I peruse the most influential evacuation solutions (that capitalize on black hole complementarity in one form or another), review their impact, and dissect their holographic connotations for Bekenstein-Hawking entropy.

The tone noticeably shifts in the second half of the chapter, where I transition in Section 3.5 to evaluating proposals on the merit of their interpretation of Bekenstein-Hawking entropy. I depend heavily on the literature concerned with guiding principles to frame the desiderata, and I argue that in order for quantum gravity approaches to move forward, we must rely on a hefty, master guiding principle to catalyze a trickle-down effect to other guiding principles. To that end, I recommend my formulation of the Universality Argument (UA), which embraces common thermodynamic and statistical mechanical behaviors across self-gravitating and non-gravitating systems alike. The most imperative function of UA is to prescribe a rubric to appraise safehouse and evacuation solutions, thereby facilitating axing either category. The interpretation of Bekenstein-Hawking entropy is thus a hypothesis to be tested against UA's determination of relevant multiply realized phenomena.

However, unlike terrestrial systems, black holes are thought to possess causal barriers, so cashing out physically salient similarities with that caveat is a nontrivial task. I contend that both camps currently falter in underwriting thermodynamic phenomenology with the statistics of Bekenstein-Hawking degrees of freedom, though for different reasons. Two preliminary criteria pitting safehouse and evacuation solutions against each other involve black holes mediating thermal contact and radiating at exactly thermal spectra. The devil is in the details of whether or not competing idealized preconditions can or should be satisfied, which is why further work needs to be done by way of a systematic comparative analysis to make either camp more convincing vis-à-vis UA.

3.2 Why Face the Phantom Music?

In this section, I aspire to leave no doubts about the circumstances under which we should tackle the black hole information loss paradox head on – epistemic dedication to black hole thermodynamics/statistical mechanics as a fruitful discipline. But before I get to that stage, here's a brief refresher of Chapter 2. Semi-classical gravity is

slyly ambiguous on how to physically interpret the mixed state describing late-time Hawking radiation – the culprit of post-evaporation information loss. To illuminate the contradictory nature of that ambiguity, I had recast the black hole information loss paradox as the paradox of phantom entanglement.

Paradox of Phantom Entanglement:

1. Late-time Hawking radiation is a global system.

2. Late-time Hawking radiation is an entangled subsystem.

Conclusion:

3. Therefore, late-time Hawking radiation is a global system and an entangled subsystem.

The first premise follows from my formulation of the general relativistic argument, which establishes that the black hole singularity destroys interior degrees of freedom so that only exterior degrees of freedom make it out to future infinity as Hawking radiation. The second premise follows from my formulation of the quantum theoretic argument, which establishes that exterior degrees of freedom eventually manifesting as Hawking radiation constitute an entangled subsystem for all time, unitarily evolving from early to late times. As is apparent, the two premises engender a contradiction. The black hole singularity has apparently annihilated the complementary interior entangled subsystem, leaving the surviving Hawking partners in the lurch to be entangled with phantoms.

Furthermore, the paradox of phantom entanglement results in information loss over the course of black hole evaporation due to a net decrease in maximal Boltzmann entropy, a measure of global information storage capacity, concurrently with a net increase in global von Neumann entropy, a measure of hidden information due to nonlocal correlations with the environment. The universe starts out with maximal Boltzmann entropy consisting of both interior and exterior degrees of freedom. It also starts out with zero von Neumann entropy, reflecting a global pure state with no external entanglement. At the final evaporation event, however, the universe's maximal Boltzmann entropy drops due to the elimination of interior degrees of freedom. Additionally, the universe's von Neumann entropy spikes due to the positive von Neumann entropy of the leftover Hawking radiation that has been conserved, reflecting a global mixed state with significant external entanglement. This conceptualization of black hole information loss aptly capitalizes on information theory for its framing, and it unveils how the designation of an externally entangled global system is an oxymoron that's physically and metaphysically unintelligible.

The bright side is that resolving the paradox of phantom entanglement is conceptually very straightforward. Late-time Hawking radiation can coherently be in only one of two mutually-exclusive states: 1) Either it's an unentangled global system, or 2) it's an entangled subsystem. But actually implementing the necessary modifications

to Hawking's framework or other derivations that have independently reproduced his result is a formidable undertaking.

So then, why bother? What's so compelling about black hole evaporation to plunge ahead in grueling, unknown territory, when it'd be perfectly sensible to dismiss the phenomenon? The spectrum of opinions regarding the answer to this question contains a continuum of nuanced philosophical positions. In order to strip the debate to its bare-bone essentials, I'm going to bifurcate the spectrum into two umbrella camps. I maintain that adherents of Camp 1 do not have high stakes in rehabilitating black hole evaporation because they reject the physical salience of black hole thermodynamics. Adherents of Camp 2, on the other hand, do have high stakes in rehabilitating black hole evaporation because they embrace the physical salience of black hole thermodynamics and its reducibility to black hole statistical mechanics.

3.2.1 Physical Salience of Black Hole Thermodynamics and Statistical Mechanics

To figure out how black holes may or may not be physically salient thermodynamic and statistical mechanical systems, let's review the basics (see also Wald 2001; Wallace 2018). Evaporating black holes radiate exactly thermal radiation at Hawking temperature T_H, which is inversely proportional to their mass M in the Schwarzschild case (setting $G = c = \hbar = 1$ in Equation 3.1):

$$T_H = \frac{1}{8\pi M}. \tag{3.1}$$

The thermality of Hawking radiation indicates a frequency distribution obeying the Planck spectrum of black body radiation, evocatively connecting black holes to black bodies. A black body emits and absorbs radiation when its microscopic constituents are perturbed, in which the frequency distribution is determined solely by temperature. Out of Clausius's Law falls thermodynamic, Bekenstein-Hawking entropy S_{BH}, famously linked to black hole area A (see Equation 3.2):

$$S_{BH} = 4\pi M^2 = \frac{A}{4}. \tag{3.2}$$

To this day, the very notion of black hole thermodynamics is fraught with controversy, and the drive to observe Hawking radiation has spurred a quest for indirect empirical confirmation. The role of analogue experimentation has prompted heated debate over critical philosophy of science issues, such as the scope of inter-type uniformity (when different systems instantiate the same properties) and the methodology behind demarcating universality classes (when different systems can be grouped together) (see Thébault 2019; Crowther et al. 2021). There's vehement disagreement over how far we can invest the formal analogy between Bekenstein-Hawking entropy and thermodynamic entropy with physical meaning, with Camp 1 suspecting the rigor of the calculations without empirical evidence and Camp 2 jumping at the opportunity for unification in quantum gravity.

Camp 1: Striking Mathematical Analogy Adherents of Camp 1 reject the physical salience of black hole thermodynamics. Their main argument is that the mathematical similarity between geometric theorems and phenomenological laws showcases a striking analogy but is merely formal in the absence of direct or indirect empirical confirmation. For example, Dougherty and Callender (2016) dispute whether the thermodynamic notion of 'equilibrium' suitably applies to black holes. 'Stationarity' is the analogue that's supposed to represent the equilibrium state, in which the metric is time-independent and invariant under time-translations. Though it's widely assumed that black holes undergoing dynamical collapse will settle into stationarity, they aver that to be in 'equilibrium with' is a critical relation between thermodynamic systems, and it's unclear whether black holes can ever be in equilibrium with each other or anything else.

The objection over the analogy between equilibrium and stationarity is related to the assertion that Hawking radiation is a kinematic effect as opposed to a dynamical one. Hawking's derivation of thermal radiation takes place in quantum field theory on curved spacetime without back-reaction. Quantum fields satisfying vacuum conditions merely ride the classical geometry of vacuum black hole solutions as opposed to being coupled. The Hawking effect comes about through state transformations (such as the Bogoliubov transformations that Hawking 1975 employed), where if quantum fields satisfy vacuum conditions in one spacetime region, they do not satisfy the same vacuum conditions in another spacetime region separated by intervening curvature. Thus, in the absence of temporal parametrization by a dynamical law, which would be the Semi-Classical Einstein Field Equation (SEFE) in this scenario, Dougherty and Callender (2016) are skeptical that Bekenstein-Hawking entropy can enforce GSL.

A more convincing formulation of their argument, which Curiel (2023a) elucidates even though he doesn't belong to Camp 1, is that progressing from Hawking radiation to black hole evaporation requires a leap of faith. As counterintuitive as it sounds, Hawking radiation doesn't automatically extract energy from its host black hole. Demonstrating energy flux requires the framework of semi-classical gravity, where classical geometry also responds to quantum matter fields. Yet the exotic energy conditions instrumental to the derivation of Hawking radiation likewise doesn't guarantee energy conservation in semi-classical gravity, a red flag for Maudlin et al. (2020). Therefore, without an independent derivation in a quantum gravity approach, black hole evaporation must be put in by hand (Curiel, 2023b).

In the absence of robust similarities between black holes and non-self-gravitating thermodynamic systems, Dougherty and Callender (2016) and Wüthrich (2019) concur that any semblance of black hole entropy must have purely epistemic roots. Wüthrich (2019) expounds the logical gap between information-theoretic (i.e., Shannon) entropy and physical entropy. Information-theoretic entropy is more general than thermodynamic entropy, which means that black hole entropy can be the former without being the latter. He also questions why Bekenstein's rationale of hidden entropy behind the horizon from the perspective of exterior observers is intimately related to surface area, which presupposes that the black hole geometry is carrying this hidden entropy, not

the matter inhabiting the interior.

The important takeaway about Camp 1 is that many of its proponents do remain open about the potential for stronger parallels between terrestrial thermodynamics and black hole thermodynamics; they just have a higher threshold for acceptance. They're acutely aware of our precarious empirical standing vis-à-vis Hawking radiation, which is the most powerful link between black holes and thermodynamic behavior (see Crowther et al. 2021), as well as our continued wait for a confirmed theory of quantum gravity that would shine some light on microscopic degrees of freedom. They also acknowledge the possibility that black hole entropy may turn out to be distinct from thermodynamic entropy, whether information-theoretic, entanglement-based, etc.

Camp 2: Unification in Quantum Gravity Adherents of Camp 2, on the flip side, accept the physical salience of black hole thermodynamics/statistical mechanics. Their confidence in black hole evaporation stems from classical clues about black hole thermodynamics in addition to quantum theoretic motivations for multiply realized statistical behaviors across weakly and strongly gravitating systems alike.

El Skaf and Palacios (2022) analyze how the study of black hole thermodynamics has gleaned substantial epistemic value from thought experiments. These hypothetical, controlled scenarios reveal inconsistencies in idealized contexts and facilitate the identification of initial assumptions that need tweaking. Along with Wallace (2018), they recount iterations of elaborate setups that expose conflicts between general relativity and the Second Law of Thermodynamics. Smoothing them over entails associating horizon area with thermodynamic entropy to prevent exterior observers from exploiting rotational or gravitational energy to routinely undermine the Second Law, perhaps by building perpetual motion machines outside a black hole. Moreover, Jacobson (1995) and Curiel (2014) characterize spacetime regions beyond causal horizons as heat sinks, claiming that spacetime physics is indeed phenomenological with a deep connection to thermodynamics.

Over and above the classical motivations for black hole thermodynamics, Hawking's derivation of thermal radiation tipped the scales in its favor. It's true that Hawking (1975) discovered a kinematic effect by splitting horizon field modes into ingoing and outgoing components and mapping the outgoing modes to those at future null infinity. But he nevertheless justified black hole evaporation without an explicit semi-classical calculation by reasoning that the inverse relation between Hawking temperature and black hole mass sufficiently controls the pace of evaporation to model as a quasi-static process until the Planck scale. By performing state transformations for different values of black hole mass, one can string together a succession of stationary states to reconstruct the dynamical evolution.

To really strengthen the physical salience of black hole thermodynamics, however, we need to demonstrate that Hawking radiation extracts energy from its black hole source, thereby reducing its mass. Wallace (2018) cites the multiplicity of (supposedly) independent derivations of a dynamical Hawking effect as theoretical evidence to collectively raise our credence in black hole evaporation. For example, solving SEFE

in asymptotically flat spacetime allows us to appeal to global energy conservation. We find that the only suitable metric (the Vaidya metric) entails the time-dependence of black hole mass, and energy conservation requires outgoing modes to be thermally-distributed at Hawking temperature.

Another derivation employs the membrane paradigm, an inter-theoretical methodology to study perturbed black hole horizons in terms of fluid mechanics and charge dissipation (see Susskind and Lindesay 2004). It solidifies the intuition of evaporation by predicting negative energy flow across the timelike stretched horizon. Furthermore, Parikh and Wilzcek (2000) formalize the heuristic of virtual particle-pair creation as a time-dependent tunneling process. When a virtual particle-pair is created close to the horizon and one partner tunnels to the other side, its energy changes sign, meaning that both particles materialize, loosely speaking, with counterbalancing energies.

The upshot of the dynamical emission of Hawking radiation is that it imbues force into GSL and allows for thermal contact, facilitating energy-mining or a black hole Carnot cycle. When coupling a black hole with a non-self-gravitating thermodynamic system, such as a photon gas reservoir, the two systems must be sufficiently far apart to neglect their mutual gravitational attraction, but nonetheless contained in a moderately sized box with reflecting walls to prevent the black hole from evaporating too quickly. Despite these contrived idealizations, exchanges of heat, amount of work performed, and changes in entropy all follow the established laws of thermodynamics, which is touted as robust theoretical evidence in favor of black hole thermodynamics naturally extending terrestrial thermodynamics (see Prunkl and Timpson 2019; Wallace 2018).

Yet according to Curiel (2023a), what's still missing in the semi-classical story is the theoretical infrastructure establishing black holes as genuine black bodies that radiate when their microscopic constituents are perturbed. In order to complete the analogy and demonstrate physical salience beyond phenomenology, adherents of Camp 2 are also responsible for putting forth a reductionist account linking emergent behavior to statistical microphysics. For that reason, they are also more heavily invested in current quantum gravity approaches, most notably string theory and loop quantum gravity. They motivate such a project by pointing to Bekenstein-Hawking entropy in standard international units (see Equation 3.3):

$$S_{BH} = \frac{c^3 A}{4G\hbar}. \tag{3.3}$$

The combination of restored constants c, the speed of light, G, Newton's gravitational constant, and $\hbar$, the reduced Planck constant, strongly hints that black holes are composite quantum gravitational systems (see Carlip 2014). Any quantum gravity approach that considers itself a serious contender better be able to not only recover Bekenstein-Hawking entropy through microstate enumeration (see Crowther 2018a), but also explain how the underlying degrees of freedom are related to horizon area. I will exposit in subsequent sections to what extent string theory and loop quantum gravity accomplish this feat, but here in the exposition, it's worthwhile to note that

they do so with prima facie unrelated ontologies, dynamics, and most importantly, radically different accounts of the role of horizon area as an entropy bound.

The important takeaway about Camp 2 is that it represents a stronger commitment to the epistemic virtues of breadth, theoretical integration, and unification. Criteria for black hole thermodynamics and statistical mechanics guide quantum gravity to grapple with and demystify currently astounding coincidences. Even though Hawking radiation knows about the black hole in quantum field theory on curved spacetime, Curiel (2023a) marvels how the classical metric somehow reciprocates that knowledge without possessing the quantum degrees of freedom to execute it when back-reaction is inserted by fiat. For Camp 2, black hole evaporation epitomizes a promissory note from quantum gravity to deliver Bekenstein-Hawking entropy such that it's "non-miraculous" (Wallace, 2020, p. 226).

3.2.2 Black Hole Information Loss is the Terrain of Camp 2

Physicists and philosophers alike have found themselves talking past each other precisely because their commitment to or rejection of presuppositions about black hole thermodynamics and statistical mechanics completely alters the parameters of the problem, which is why many of them have been bickering over the stakes of information loss. It has become apparent to me, drawing inspiration from Dougherty and Callender (2016)) and Wallace (2020), that Camp 1 has very low stakes in black hole information loss and the paradox of phantom entanglement. Since skeptics of black hole evaporation lack incentives to rescue Hawking's framework, they have no business either proposing or weighing in on prospective solutions.

Camp 2 then is the primary driver of the black hole information loss discourse. Its advocates make the most of black hole evaporation as an opportunity to unify thermodynamics and statistical mechanics among the most disparate of systems. Unification in this sense occurs through coarse-grained homogenization, in which Batterman (2000) explains that common macroscopic behaviors, i.e., multiply realized phenomena, emerge from homogenizing what are actually microscopically heterogeneous systems. The sentiment is that black holes realize behaviors common to all thermodynamic statistical systems.

> Specifically, we assume that the origin of the thermodynamic behavior of the black hole is the coarse graining of a large, complex, ergodic, but conventionally quantum mechanical system (Susskind et al., 1993, p. 3746).

I will explore this motivation in greater detail in Section 3.5.5. But for now, in order for Camp 2 to even take off, black hole evaporation must be recovered in a suitable semi-classical framework, though of course, suitability is determined with some retroactive exegesis of the required modifications to Hawking's derivation, as I will analyze in Section 3.5.3. That said, because thermodynamics is inherently a spatiotemporal discipline, the desideratum of continuity between black holes and pedestrian systems calls for recovering black hole thermodynamics in an effectively

classical spacetime. Entropy is quantified through spatial boundary conditions, like volume or area. Moreover, ensuring the (statistically valid) global non-decrease of entropy, $S(t)$, in the Second Law hinges on temporal boundary conditions grounding the arrow of time (see Equation 3.4 and Albert 2000):

$$\frac{dS_{global}}{dt} \geq 0; S(t_0) \leq S(t). \tag{3.4}$$

Therefore, the protection of GSL calls for exhaustively carving up an evaporation spacetime into simultaneity slices. That's why I remained steadfast in previous chapters and refused to entertain the prima facie pathological spatiotemporal structure of Hawking's prenrose diagram (see Figure 3.1) as a decisive refutation against even posing the black hole information loss paradox. The vast majority of proposals support Hawking's expectation that black holes evaporate in sub-Planckian regimes.

> **Black Hole Evaporation Conjecture (BHEC)**: Black holes evaporate at least until reaching Planck mass.

Otherwise, the trust in black hole evaporation simpliciter falters, defeating the purpose of this camp (see Preskill 1992).

3.2.3 The Exorcism: Modifying Hawking's Framework

The foresight of the paradox of phantom entanglement is that fixing the contradiction and choosing between late-time Hawking radiation as a global system versus an entangled subsystem uncovers how Camp 2 quickly disintegrates into rival branches, where the main point of contention is whether or not Bekenstein-Hawking entropy accounts for the totality of black hole degrees of freedom or solely those of the horizon.

Before we get ahead of ourselves, we need to figure out what fixing the contradiction entails. The precise modifications to Hawking's original framework conducive to exorcising entangled phantoms will depend on the details of the proposal. However, it's pretty glaring that minimally, the spacetime structure of Hawking's Penrose diagram needs to be updated.[1] Resolving the paradox of phantom entanglement involves: 1) transforming the black hole from an annihilator into a *safehouse* or 2) facilitating the *evacuation* of trapped degrees of freedom. Implementing these strategies requires a better understanding of the physics of three critical regions that are highlighted in Figure 3.1.[2]

First, the spacetime region in the vicinity of the singularity, demarcated in yellow, approaches infinite curvature. Singularity resolution would effectively "plug the

[1] Maudlin (2017) proposes foliating Hawking's Penrose diagram differently instead of changing the spacetime structure. I disagree that his recommended foliation is innocuous without invoking auxiliary structure, but I do not have the scope to engage further here and am postponing a more detailed discussion to future work.

[2] I'd like to thank Christian Wüthrich for discussions on modifying an evaporation Penrose diagram.

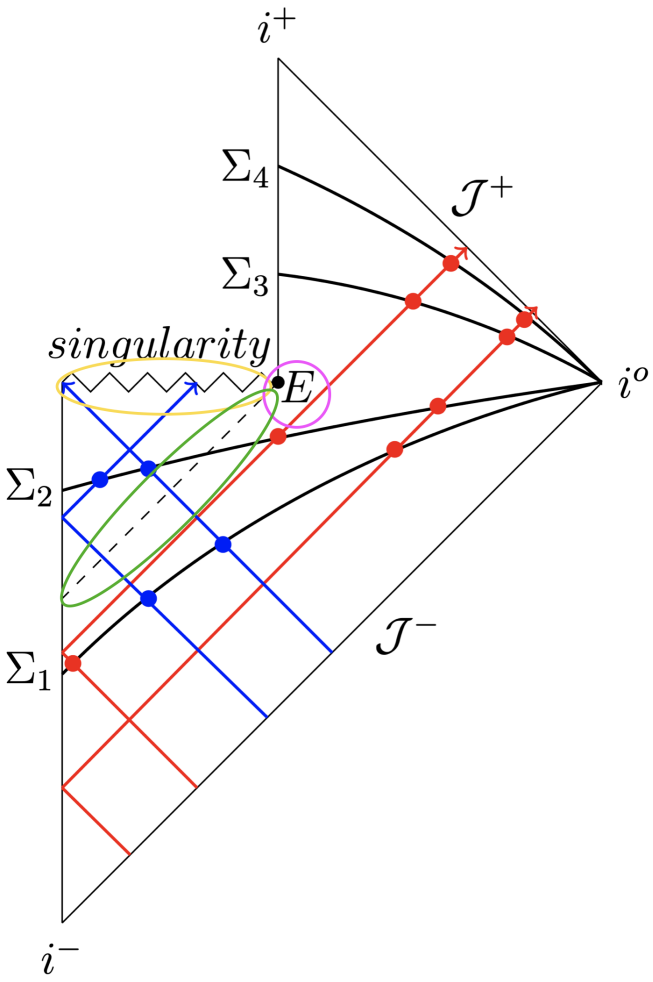

Figure 3.1: Domains of Quantum Gravity
The pathological structure of an evaporation spacetime encompasses three critical regions where the need for a theory of quantum gravity becomes evident. The pink region around E indicates ambiguity in late stages of black hole evaporation. The yellow region prior to the singularity signifies a curvature blow-up. And the green region in the vicinity of the event horizon insinuates violations of the no-hair theorem.

information leak" (Earman, 1996, p. 634) and independently serves as a guiding principle for the development of quantum gravity approaches (Crowther, 2018a). Second, the semi-classical calculation is only valid until the black hole reaches Planck mass, around 10^{-33} g (Hawking, 1975), so the process of late-stage evaporation and the fate of the final evaporation event E are unknown. This ambiguity is indicated in the pink region. Both of these regions are candidate safehouse locations for the complementary subsystem entangled with late-time Hawking radiation.

Third, the region just outside the event horizon (depicted in green) may no longer be constrained by the no-hair theorem and could transmit information to outgoing Hawking radiation. It involves novel physics, perhaps from highly excited transplanckian modes (see Susskind and Thorlacius 1994; t'Hooft 1996); Almheiri et al. 2013),

or alternatively, "soft hairs" that don't interact energetically (see Hossenfelder 2012; Hawking et al. 2016). The event horizon and immediate surroundings are therefore a candidate location for an evacuation mechanism.

The majority of well-known proposals in the physics literature can thus roughly be classified into two, mutually exclusive families of views: 1) safehouse solutions and 2) evacuation solutions. As a pedantic note, 'solutions' is unquestionably a conditional label. What all of these proposals have in common are treatments from prototypical quantum gravity approaches to handle the physics of Planckian regimes, and as we will see in the coming sections, Bekenstein-Hawking entropy is the conceptual core of all iterations of the black hole information loss paradox. I argue that safehouse solutions insist on a black hole's internal information storage capacity being unbounded in the infinite limit, whereas evacuation solutions insist on it being bounded by Bekenstein-Hawking entropy. For the minority, renegade proposals aspiring to combine safehouse and evacuation dynamics, I'm going to categorize them predominantly as safehouse solutions because of their interpretation of ever-growing internal storage capacity.

3.3 Safehouse Solutions: Black Hole Entropy is Causal

Historically, the immediate, subliminal reaction to the paradox of phantom entanglement has been to reject the conclusion that late-time Hawking radiation is a global system (see e.g., Giddings 1992; Polchinski and Strominger 1994). The quantum theoretic argument prevails over the general relativistic argument, which deems late-time Hawking radiation as an entangled subsystem. Of course, we must now rely on a quantum gravity approach to appropriately modify Hawking's semi-classical framework and ensure the persistence of a black hole remnant to host the complementary entangled subsystem.

3.3.1 Rundown of Black Hole Remnants

The weakest link in the general relativistic argument is the premise establishing the annihilation of black hole interior degrees of freedom by the singularity. Any solution transforming the interior into a safehouse has several options to deny the problematic premise. The first, most obvious option is to get rid of the singularity and extend the interior spacetime, a minimal requirement if the event horizon disappears. The second option is to halt evaporation when the event horizon is extremely small, i.e., of Planck scale. Since the radius of the event horizon depends on black hole mass, all safehouse solutions can be characterized either as massless or massive remnants of various types.

Causally-Connected Massless Remnants: Hawking (1975) postulated that when a black hole evaporates completely, the surrounding metric resembles flat spacetime. Removing the singularity opens the door to converting the final evaporation event E

from a topological discontinuity as seen in Figure 3.1 to a continuous point on the manifold with the Minkowski metric. As Unruh and Wald (2017) remark, the interior degrees of freedom that would've otherwise been annihilated by the singularity must now be exposed. After the disappearance of the event horizon and in the absence of curvature, the once trapped complementary entangled subsystem becomes liberated as a causally-connected, massless remnant.

In the proposal of Perez (2017), for instance, late-time Hawking radiation is entangled with Planck-sized "atoms of geometry" constituting the underlying spacetime. Any continuous metric is but a coarse-grained approximation of the discrete granular structure of this quantum gravitational substrate. In the paradox of phantom entanglement, we couldn't find the complementary entangled subsystem *in* the post-evaporation spacetime because it *was* the post-evaporation spacetime, where Planckian degrees of freedom were traced out of the effective, Minkowski description. In fact, semi-classical gravity inevitably traces out Planckian degrees of freedom by virtue of treating the metric classically. So, a collective remnant is always hidden in the background throughout black hole evaporation. Its macroscopic nature, such as its mass and causal relationship to Hawking radiation, is determined by the emergent metric.

Causally-Disconnected Massless Remnants: In other proposals, however, black holes leave massless remnants that become causally disconnected from late-time Hawking radiation, more popularly known as baby universes. Hsu (2007) explains that the yellow-region of high-curvature in Figure 3.1 is dominated by quantum gravitational tunneling events which can lead to abrupt topology change. Quantities that are classically conserved in the parent universe would actually prevent quantum gravitational tunneling, such as mass (energy), angular momentum, and charge, which is why they're radiated away. Therefore, the baby universe that pinches off does not violate energy conservation in the parent universe.

In fact, Hossenfelder and Smolin (2010) corroborate abrupt topology change in their tweaked Penrose diagram, which portrays the final evaporation event E as the discontinuous pinch-off point from the parent to baby universe. In their Penrose diagram, the singularity is removed and the black hole interior extended to a compact boundary that's spacelike separated from future timelike infinity. Degrees of freedom entering the black-hole-turned-baby-universe now measure infinite proper time just like their counterparts in the parent universe, thus avoiding annihilation. Therefore, the complementary subsystem entangled with late-time Hawking radiation remains in baby universe. Due to the lack of continuous spacelike curves connecting the parent and baby universes, a non-Cauchy surface like Σ_3 is embedded in a disconnected Cauchy surface, whose associated pure state encodes nonlocal correlations.

Stable Massive Remnants: On the other end of the remnant proposal spectrum, some out-of-the-blue mechanism halts the evaporation process so that we're left with a massive remnant. As Chen et al. (2015) concede, no massive remnant proposal has worked out the dynamics, with speculations ranging from adding higher curvature

gravitational terms to the Einstein-Hilbert action to adding matter fields in order to stabilize non-evaporating, electrically charged (i.e., extremal) black holes under perturbations. Nonetheless, they contemplate insights from general quantum gravitational principles, most notably, a generalized Heisenberg uncertainty principle between position spread out over a minimal length and momentum. Since this minimal length is defined in terms of Planck mass, and Hawking temperature depends on black hole mass, the generalized uncertainty principle affects late stages of evaporation.

A useful relationship between mass and temperature in general is specific heat, which sets the threshold for how much energy it takes to raise a system's temperature by an incremental amount. Black holes actually have negative specific heat – their temperature increases when they lose mass-energy – normally resulting in runaway evaporation (see Wallace 2018). But the generalized uncertainty principle guarantees that when a black hole reaches Planck mass, its specific heat becomes infinitely negative, or equivalently zero, thereby halting evaporation (see Chen et al. 2015). A stable massive remnant then serves indefinitely as a safehouse for interior degrees of freedom entangled with late-time Hawking radiation.

Decaying Massive Remnants: Decaying remnants, in contrast, serve as safehouses for far longer than the present age of the universe but eventually evacuate all interior degrees of freedom to purify the final state. Going off the toy model of Hawking pairs, the trapped entangled partners make their way across the horizon and rejoin their other halves, thus changing sign from negative to positive-energy. In such proposals, Hawking radiation is entangled with a black hole remnant before the second part of the dynamical story, after which information is returned to the exterior (see Chen et al. 2015).

For example, Bianchi et al. (2018) investigate how black holes could tunnel into white holes, their time-reversed counterparts. While this process is suppressed classically, such events dominate at the Planck scale. Therefore, the suggestion is that the second part of the dynamical story of black hole evaporation involves a white hole remnant, in which entry, not escape, is forbidden by the speed-of-light barrier. The path of least resistance now for interior degrees of freedom is out to future infinity, and for reasons that will become clearer later, it's vital that they leave extremely slowly as "soft", low-energy quanta.

3.3.2 Interim Challenges and Responses: Too Little Energy for Too Much Information

The primary critique of safehouse solutions is that Planck-mass black holes are considered too small, possessing too little energy to either release or store the vast amount of information associated with the enormous complementary entangled subsystem. Hawking (1976) himself didn't find remnants to be physically feasible.

> [I]nformation like baryon number requires energy and there is simply not

> enough energy available in the final stages of the evaporation. To carry the large amount of information needed would require the emission in the final stages of about the same number of particles as had already been emitted in the quasistationary phase (p. 2472).

Hawking's argument is speciously attractive, which is why decaying remnants have been neglected in the discourse until they've made a pretty recent comeback. He was operating under the assumption that remnants of any kind were untenable, probably stemming from the expectation that Planck-mass black holes have negative specific heat, so runaway evaporation would mandate immensely fast evacuation of interior degrees of freedom to return information to the exterior.

Moreover, according to Bekenstein (1981), "[F]ast information is energy expensive" for non-equivalent modes (p. 625). Even Wald (2019), who was once an avid supporter of causally-connected massless remnants (see Unruh and Wald 2017), claims that in four-dimensional spacetime, late-time radiation must be entangled with degrees of freedom that have Planckian energies. This finding is flatly incompatible with the Minkowski metric and vacuum degeneracies storing information.

However, if the emission of the complementary entangled subsystem as 'Hawking-like' or non-Hawking radiation takes place extremely slowly in Planckian regimes, akin to one quantum at a time, the same number of particles as had been emitted in the semi-classical phase could also be emitted in the quantum gravitational phase (Chen et al., 2015). The turtle-paced decay would thus render the black hole remnant quasi-stable, and by the end, information would be restored energetically cheaply. Spelling out the details of decelerating the rate of emission is imperative for the viability of safehouse solutions and a work in progress (see Bianchi et al. 2018).

Furthermore, while Planckian remnants are stable, whether long-lived or permanent, a related concern is that they're practically point particles from the perspective of the exterior, yet they can hide endlessly rich internal structure. The same concern is pertinent to baby universes and causally-connected massless remnants, which have lost every ounce of gravitational energy to Hawking radiation. Skeptics are unconvinced that structure/information can be decoupled from energy, similar to what Mann (2015) expressed in a quote earlier.

The most common rebuttal alludes to violations of certain energy conditions in the derivation of black hole evaporation, thereby questioning the sanctity of energy conservation in semi-classical gravity (see Chen et al. 2015; Maudlin et al. 2020). Since that counterargument runs afoul of black hole evaporation in the first place (perhaps intentionally by Maudlin et al. 2020), a more powerful loophole is that energy is a frame and coordinate-dependent variant. Therefore, a remnant's energy as reported by a distant observer at asymptotically flat infinity (i.e., its ADM mass) is much lower than that which is reported by an observer inside, where the rich structure resides.

Ha (2003) corroborates this thinking by contemplating the net energy of a black hole with a test mass inside, as per an observer at asymptotically flat infinity. When the negative gravitational potential energy between the black hole and test mass is

factored in, the black hole's mass is reduced by that amount. He conceptualizes the negative gravitational potential energy as positive kinetic energy waiting to be extracted from the black hole, which is precisely how Hawking radiation is emitted. A distant observer thereby concludes that positive-energy radiation sources its fuel from the negative gravitational potential generated by partners behind the horizon, though some of that fuel is used up to escape the potential well (i.e., in red-shifting). With enough test masses behind the horizon, a distant observer reports the black hole's net energy to decrease commensurately, but as Hsu (2007) confirms, the combined masses and rich structure linger for an interior observer.

Even if we buy into the energy-efficient, structure-forming capabilities of the negative gravitational potential, the rich internal structure of remnants becomes worrisome when they quasi-stabilize or permanently stabilize at the Planck mass. If evaporation time scales permit them to mingle with ordinary matter, the fear is that they would upend all known physics due to wildly exotic and unstable interactions (see Polchinski 2017). This situation is even more unsavory if there could be infinitely many species of remnants, given that they form and store information from arbitrary initial states. It would then be entropically favorable for the universe to produce more and more remnants (see Preskill 1992).

For these and other reasons, many physicists had abandoned the remnant route. However, there's been a resurgence of interest in safehouse solutions due to a critical reassessment of whether the initial objections coming from effective field theory fairly apply to Planck-scale physics, and most notably, because of developments in interpreting Bekenstein-Hawking entropy as entanglement entropy (Chen et al., 2015).

3.3.3 Bekenstein-Hawking Entropy: Horizon States

Now that we have walked through and gotten the gist of several safehouse solutions, we're in a better position to infer what they imply about the interpretation of Bekenstein-Hawking entropy. As a quick but pertinent digression, I'd like to clarify why I did not group causally-connected massless remnants and decaying massive remnants with evacuation solutions, even though they maintain that all black hole degrees of freedom eventually become exposed to the exterior spacetime and constitute the final global system. I believe it's helpful to view them as bait-and-switch proposals that have more in common with other safehouse solutions, all of which compel Hawking radiation to remain maximally entangled with the interior until the black hole reaches Planck mass. This entails that Bekenstein-Hawking entropy is not synonymous with overall black hole entropy and is but a fraction of its total information storage capacity.

As I touched upon in Chapter 2, the first requirement for a black hole remnant is to purify late-time Hawking radiation, which implies that its Boltzmann and von Neumann entropies must match. To gauge how large these quantities are, let's carry out a rough calculation employing the toy model of Hawking pairs, in which positive and negative-energy quanta are monogamously entangled. The reduced density matrix of a positive-energy partner has two eigenvalues, p_1 and p_2. Moreover, monogamous entan-

glement leads to the inference of a uniform probability distribution, so each eigenvalue is $\frac{1}{2}$ and weights an independent degree of freedom. Therefore, a positive-energy member of a Hawking pair has a two-dimensional Hilbert space with Boltzmann entropy of $\ln 2$. In fact, its von Neumann entropy is also $\ln 2$ (where $S_{VN} = -\sum p_i \ln p_i$), reflecting saturated entanglement with its negative-energy partner.[3] For n Hawking pairs, the Boltzmann and von Neumann entropies of the collection of positive-energy quanta comprising late-time Hawking radiation equal $n \ln 2$. The same reasoning applies to the collection of negative-energy quanta living inside a black hole remnant.

To further get a sense of the information storage capacity required of the complementary entangled subsystem, let's put n, the number of either positive or negative-energy quanta, into perspective. Energy conservation entails that late-time Hawking radiation compensates for the initial black hole mass M for massless remnants, or at least $M - m_p$ for massive remnants, in which $m_p \sim 10^{-8}$ kg is the Planck mass. According to Mann (2015), if we divide M by the average energy emitted per quantum that's a tiny fraction of the Planck mass, $E_q \sim \frac{m_p^2}{M}$, we can solve for n. After minor rearranging, it turns out that $n \sim (\frac{M}{m_p})^2$. For a solar mass black hole of $M \sim 10^{30}$ kg, $n \sim 10^{76}$. Contrast this quantity with the Bekenstein-Hawking entropy of the initial black hole (see Equation 3.3). Again for a solar mass black hole, $S_{BH} \sim 10^{76}$, of the same order of magnitude as n. This comparison provocatively suggests that the entropy of a post-evaporation remnant that's lost most of its mass and surface area must be at least as much as that of a pre-evaporation black hole with all of its mass and surface area intact.

As of yet, we don't know much about the statistical or ontological foundations of Bekenstein-Hawking entropy, but if we take safehouse solutions at face value, we're committed to saying that S_{BH} can't be nearly large enough to account for the information storage capacity inside the black hole. Note that the production of Hawking radiation increases over time, as more negative-energy quanta cross the event horizon, which is depicted in Figure 3.1 in the evolution between Σ_1 and Σ_2. But because horizon area is also decreasing during this period, there's a time – the Page time – at which the black hole's Bekenstein-Hawking entropy becomes too small for it to wholly couple with Hawking radiation (see Page 1993b). Figure 3.2 summarizes this analysis by graphing the behavior of various entropy curves over time.

All of the curves are dashed instead of continuous to represent incremental changes in entropy with the production of discrete Hawking pairs. S_{BH}, the blue dashed curve, is the Bekenstein-Hawking entropy of a black hole subsystem, understood as thermodynamic entropy though the statistical underpinnings are unknown. It's maximized when the black hole hasn't yet begun evaporating and is at its largest mass/surface area. Over time, the rate of Bekenstein-Hawking entropy decrease accelerates to capture the runaway evaporation effect of the black hole's negative specific heat.

S_{rad}, the red dashed curve, is the thermal entropy of Hawking radiation. It's min-

[3]The goal is to provide a sense of scale, so I'm setting aside the non-uniform emission rates of a thermal distribution, which would also be reflected in the degree of entanglement.

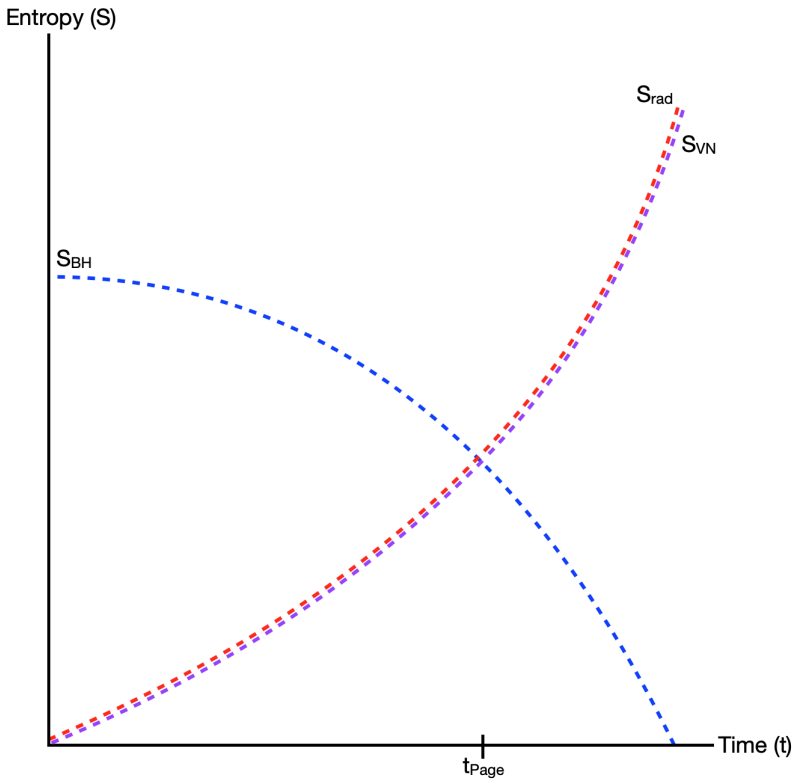

Figure 3.2: Safehouse Solutions: The Hawking Curve
The black hole's Bekenstein-Hawking entropy, which is proportional to surface area, decreases over time in discrete steps with the absorption of negative-energy quanta, as shown by the blue dashed curve. The exterior radiation's thermal entropy increases over time in discrete steps with the emission of positive-energy quanta, as shown by the red dashed curve. The von Neumann entropy between the black hole and exterior radiation increases over time in discrete steps with the production of more Hawking pairs, as shown by the purple dashed curve. All three quantities are equal at the Page time.

imized prior to the separation of Hawking pairs by the event horizon, and the rate of increase accelerates for the same reason as before – to capture the runaway evaporation effect of the black hole's negative specific heat. It's maximized when Hawking radiation stops being produced, at which point its thermal entropy far exceeds the black hole's initial Bekenstein-Hawking entropy. S_{VN}, the purple dashed curve, is the entanglement entropy of Hawking radiation that arises by tracing out negative-energy degrees of freedom behind the event horizon at any given moment. It tracks S_{rad} because the Hawking radiation's thermal and reduced density matrices coincide.

Given that I went to extensive lengths in Chapter 2 to argue that the von Neumann entropy of positive-energy quanta is conserved throughout black hole evaporation, let me pause and clarify how that's consistent with S_{VN} increasing over time. If we fix the referent system as the positive-energy quanta even before they behave as Hawking radiation, i.e., on Σ_1, then entanglement is most certainly preserved under unitary evolution. However, Figure 3.2 does not fix referent system. Rather, it fixes the

boundary – the event horizon – that bipartitions the spacetime into the black hole (surface-plus-interior) and everywhere else. The degrees of freedom contained in these complementary regions are not conserved. As more negative-energy quanta traverse the event horizon, more positive-energy quanta behave as Hawking radiation, which accounts for the increase in entanglement across the boundary, S_{VN}.

The insight of Figure 3.2 is illustrating the critical transition at the Page time when the thermal entropy of Hawking radiation depletes the information storage capacity of the Bekenstein-Hawking subsystem. Afterwards, because $S_{rad} = S_{VN} > S_{BH}$, the deep interior most definitely has to compensate (if it hasn't already been doing so). What's telling is that Bekenstein-Hawking entropy decreases as the black hole evaporates, so it must be divorced from the interior complementary entangled subsystem whose entropy increases and surpasses it. This decoupling is brought into sharper relief when we realize that safehouse solutions compel the interior to have infinite information storage capacity given an infinite lifespan.

Page (1995) sketches a thought experiment in which we feed a macroscopic black hole with just enough mass-energy to counteract the rate at which it evaporates, such that its size stays stable. Because Hawking radiation is always maximally entangled with the interior when a black hole is larger than the Planck mass, the continual feeding grows the entanglement. And since we're adding degrees of freedom to the black hole, the number of internal states is also increasing. Therefore, as Polchinski (2017) emphasizes, it seems like a black hole of constant mass and surface area can harbor an unbounded number of internal states. However, I wish to take this thought experiment even further. It's not just that a black hole of constant mass and surface *can* harbor an unbounded number of internal states. For consistency's sake with the modus operandus of safehouse solutions, it *must*.

Let's imagine that the interior storage capacity of a black hole is finite and bounded, though still larger than its Bekenstein-Hawking entropy. Giddings (1992) puts forth exactly this proposal and stipulates an entropy bound as a multiple of Planck volume that can fit in a spatial region. He also conjectures that a black hole's volume grows with its mass. If we continuously feed such a black hole to prevent it from evaporating, then its internal information storage capacity will eventually become saturated. Feeding it after that time might cause the black hole to get bigger and radiate less energy, but then there's no stopping it from becoming infinitely massive, which contradicts the premise of finite storage capacity for evaporation to even take place. In the absence of intervention, a black hole with finite volume can evaporate to its entanglement saturation point while still macroscopic and then simply stop radiating, which Preskill (1992) dislikes for contradicting BHEC.

Alternatively, feeding the black hole after the saturation point might cause it to radiate degrees of freedom that are no longer entangled with the interior. Giddings (1992) concedes that nonlocal effects take over, which is what safehouse solutions find unacceptable for macroscopic black holes and is their grueling contention against evacuation solutions. Black holes and their remnant descendants are thus reminiscent of Wheeler's bag-of-gold spacetimes with constricting throats and unrestrained spatial

volume (see Marolf 2009). For example, an expanding universe (with an FLRW metric) can be embedded by a wormhole (with an Einstein-Rosen metric) into a Schwarzschild black hole of *any size* and contain *arbitrarily many states*.

Consequently, safehouse solutions deem Bekenstein-Hawking entropy to be a fraction of total black hole entropy, revealing that the black hole itself is a huge, complex system partitioned into subsystems. And as far as I can tell (given that it has slipped under the radar in the literature), the fraction of Bekenstein-Hawking degrees of freedom is of measure zero because it's finite out of an infinite sea of degrees of freedom.

That begs the question: Which black hole subsystem does Bekenstein-Hawking entropy pertain to? Safehouse solutions have a very simple answer. Because Bekenstein-Hawking entropy is proportional to surface area, it must apply only to horizon degrees of freedom. Advocates aver that solely the surface system mediates thermal contact with Hawking radiation (see Jacobson et al. 2005). The presence of an event horizon actually provides a physically preferred boundary to trace out degrees of freedom that can't contribute to a black hole's thermodynamic behavior (see Sorkin 1997; Sorkin 2011). Carving along nature's joints thus elucidates the meaning of Bekenstein-Hawking entropy. It's an information measure over those black hole degrees of freedom that are causally efficacious to the exterior. It follows then that the primary justification for safehouse solutions is what I frame as the Causality Argument (CA).

> **Causality Argument (CA)**: Black holes' thermodynamic behavior arises from interactions with external systems. Since the event horizon is a causal barrier, only surface states influencing outside regions of spacetime underlie Bekenstein-Hawking entropy. Non-thermodynamic black hole entropy is unbounded.

As Rovelli (2019) explains, for most ordinary systems, thermodynamic entropy is an information measure over all degrees of freedom available to equilibrate under putative macroscopic constraints. But for black holes, there are many more degrees of freedom available to become entangled without equilibrating – those of the interior that are causally inaccessible. Black holes are special because their thermodynamic, Bekenstein-Hawking entropy does not bound their von Neumann entropy from above. Entanglement is nevertheless still limited by total Boltzmann entropy without having to import a thermodynamic connotation. After all, we don't want to relapse to the paradox of phantom entanglement with more entanglement relations than relata to anchor them.

Based on CA, Sorkin (1997) identifies three approaches that attempt to secure the equality between S_{BH} and the degeneracy of horizon states: 1) horizon geometries, 2) micro-constituents of the quantum gravitational "substratum", and/or 3) entanglement at extremely short distances across the horizon. Since CA traces out interior degrees of freedom, it insinuates that thermodynamic, statistical, and entanglement entropy are unified in the presence of a horizon. Indeed, the claim to fame for safehouse solutions is the trifecta of Bekenstein-Hawking entropy, profoundly combining aspects of all three candidates. Consequently, the entanglement entropy of Hawking

radiation, S_{VN}, must be distinct from S_{BH}, which can only be an information measure over near-horizon modes straddling the causal boundary. As such, S_{VN} exhibits growing entanglement with black hole degrees of freedom that are not implicated in S_{BH}, namely those of the deep interior.

The nature of Bekenstein-Hawking degrees of freedom is going to depend on the theory, though physicists working on safehouse solutions tend to be heavily influenced by loop quantum gravity. Rovelli (1996) pioneered calculations within loop quantum gravity enumerating surface microstates that turned out to be proportional to horizon area, and Ashtekar et al. (1998) recovered the Bekenstein-Hawking factor of one-quarter by fine-tuning the free Immirzi parameter. So, preferred ideas about the underlying substratum include spacetime quanta and spin networks. But regardless of how Bekenstein-Hawking entropy is cashed out ontologically, its confinement to horizon states insinuates that the evolution of the surface system is decoupled from the evolution of the interior. And for causally-disconnected massless remnants, the identity of a black hole is independent of the existence of a horizon altogether seeing as the interior persists long after the evaporation of the surface system.

3.4 Evacuation Solutions: Black Hole Entropy is Holographic

Though safehouse solutions may have pervaded the discourse in the early days and recently enjoyed a renaissance, the advent of bulk/boundary dualities have long since turned the tide in favor of evacuation solutions. Evacuation solutions reject the conclusion that late-time Hawking radiation is an entangled subsystem. Far from the general relativistic argument winning out over the quantum theoretic argument, evacuation solutions are more ambitious in scope and deny the validity of both arguments. In essence, they rely on quantum gravity corrections to supplant Hawking's calculation well before an evaporating black hole shrinks down to Planck mass, leading to substantial deviations from general relativistic and quantum field theoretic predictions in what were thought to be applicable domains (see Wallace 2020).

3.4.1 Rundown of Hairy Horizons

Evacuation solutions agree with safehouse solutions that singularity resolution is a minimal requirement to resolve the paradox of phantom entanglement. They aim for late-time Hawking radiation to be a global system, not as a surviving subset of degrees of freedom as the general relativistic argument commands, but as the totality of degrees of freedom. Consequently, anything trapped in the black hole interior can't be annihilated – it must escape. Bluntly enacting this condition introduces superluminal dynamics.

To mitigate undermining the black hole horizon, evacuation solutions end up rejecting the quantum theoretic argument as well, which implicitly attributes the discontin-

uous decomposition of the vacuum into independent positive and negative-frequency modes to the causal boundary posed by the event horizon (Hawking, 1975). All evacuation solutions add quantum gravitational structure to the black hole surface system in order to evade and/or relax the causal boundary, thereby altering the behavior of fields at the effective level and precluding the clean division of Σ_2 in Figure 3.1 into the entangled subsystems of Σ_{2+} and Σ_{2-}. This strategy endeavors to get information out with finesse by undoing the no-hair theorem and bypassing the speed-of-light-barrier in a regime where novel physics is fair game.

Stretched Horizon and Black Hole Complementarity: Susskind et al. (1993) deploy the membrane paradigm, an inter-theoretical methodology to study perturbed black hole horizons in terms of fluid mechanics and charge dissipation, to substitute the black hole interior. They take advantage of complementary descriptions across reference frames – the adiabatic vacuum for inertial observers entering a black hole versus Hawking radiation for stationary observers hovering outside it – to defend that stationary observers are entitled to impose boundary conditions at the event horizon that excise the interior. After all, no exterior observer ever detects any infalling object cross the horizon due to infinite time dilation. By invoking a timelike boundary of radius that's one Planck length larger than the event horizon, we can track the evolution on this "stretched horizon" that behaves like a viscous, conducting membrane. The effective dynamics belong to conformal field theory, the boundary description of a bulk/boundary duality (see Susskind and Lindesay 2004; Chatterjee et al. 2012; Matsuo 2021).

The stretched horizon relegates highly excited Planckian modes to a quantum gravitational regime and acts as a buffer to prevent degrees of freedom from getting trapped in a causally inaccessible spacetime region. By fiat, it serves as an impenetrable, flammable barrier that absorbs, thermalizes, and re-radiates ingoing degrees of freedom as outgoing 'Hawking-like' radiation. It's worth pointing out that the stretched horizon must be an adequate scrambler. In general relativity, the scrambling time of a black hole marks the duration at asymptotically flat infinity for collapsing matter to alter the metric and elicit an event horizon. Scrambling times are also associated with subsequent perturbations to the black hole, after which, the no-hair theorem holds as a coarse-grained smearing of fine-grained details. So, for the stretched horizon to thermalize and equilibrate with incoming matter in the classically allotted time, it must involve nonlocal dynamics. The emission of radiation, however, remains local without a causal barrier obstructing its escape (see Lowe et al. 1995; Susskind 2012; Susskind 2013).

Hard/Soft Gravitational Hairs: As opposed to the membrane paradigm which approximates quantum gravitational effects in the context of bald black hole metrics, other complementarity-inspired proposals ascribe horizon hairs to gravitational back-reactions. Discussions in the 1990s centered around strong interactions near the horizon between blue-shifted ingoing and outgoing modes that would imprint upon

the black hole's geometry as well as late-time Hawking radiation. Once again, opposing descriptions across reference frames must be reconciled, where the black hole horizon and interior represent smooth regions of spacetime for infalling observers but never emerge classically for stationary exterior observers (see Kiem et al. 1995; t'Hooft 1996).

Contemporary proposals run with the idea of hairy black hole metrics, but they suspect that most of the information about the interior is not actually encoded in the horizon's 'hard hairs', which record deviations in mass/energy and curvature coming from highly excited Planckian modes. Whereas quantum gravity tends to be conflated with high-energy, UV-physics, Hawking et al. (2016) draw upon advances in low-energy, IR-physics to hypothesize that transplanckian modes of wavelengths shorter than the Planck length are unexcited, thereby creating "soft hairs", e.g., zero-energy gravitons, that substantively thicken a black hole's tresses. Soft hairs imbue Hawking radiation with information about the interior without changing its energy spectrum, making the vacuum degenerate and information-laden. Calmet et al. (2022) build upon this conjecture and demonstrate a degeneracy of graviton states for fixed horizon area capturing the black hole's internal composition. Hawking radiation produced at the horizon thus retains the interior's signature through these soft gravitons.

ER=EPR: All of the proposals thus far seek to endow the horizon with more information than it has classically to postpone detailing the inevitable – the superluminal escape of interior degrees of freedom. Yet, a classical resource to potentially bypass the speed-of-light-barrier that has recently garnered attention is the Einstein-Rosen bridge, colloquially known as a wormhole. Maldacena and Susskind (2013) spot intriguing similarities between wormholes, the classical loophole to locality, and entanglement, the quantum loophole to locality. They muse that by taking many particles in Eintsein-Rosen-Podolsky states (i.e., Bell states), separating the maximally-entangled partners, and compressing each half into a black hole, we can create two maximally entangled black holes.

They posit that this entanglement is tantamount to a connecting wormhole, leading to the famous conjecture that ER=EPR. Although Einstein-Rosen bridges are not readily traversable classically, the right quantum operations could get degrees of freedom out from one end to the other. ER=EPR thus hints at a mechanism for entangled, trapped degrees of freedom to reunite with Hawking radiation. In fact, Jusufi et al. (2023) argue that Hawking pairs epitomize the geometric realization of quantum entanglement because negative energy is required to keep a wormhole throat open, which is precisely the advantage of negative-energy partners. If an Einstein-Rosen bridge exists for every Hawking pair, the multiplicity of wormholes would affect the metric of the intervening spacetime, becoming another candidate for horizon hairs.

Entanglement Wedge Reconstruction: Penington (2020), Almheiri et al. (2020), and Engelhardt and Folkestad (2022) formally implement ER=EPR and black hole complementarity through entanglement wedge reconstruction, a methodology devel-

oped by Ryu and Takayanagi (2006) and Engelhardt and Wall (2015) to relate observables in bulk/boundary dualities. Taking some liberties here with the physical intuition for black hole evaporation, the black hole interior is the bulk and asymptotically flat infinity containing Hawking radiation is the boundary.[4] An entanglement wedge is a bounded spacetime region in the bulk whose observables can be reconstructed from the entanglement structure within its boundary. They show that during the first half of evaporation (roughly), there's no entanglement wedge inside the black hole. Interior degrees of freedom cannot be reconstructed from late-time Hawking radiation because exterior degrees of freedom aren't entangled among themselves. All positive-energy modes are entangled with negative-energy modes across the horizon, which aligns with Hawking's derivation.

During the remaining period, however, there is a growing entanglement wedge gradually encompassing the interior, indicating the increased translatability of interior degrees of freedom in terms of entanglement within late-time Hawking radiation. This result has several implications. First, entanglement among exterior degrees of freedom entails that the entangled partners of early Hawking radiation have escaped, presumably through wormholes. Second, since interior degrees of freedom are reconstructed from exterior degrees of freedom entangled among themselves, this entanglement structure enriches the classical black hole and violates the no-hair theorem. It's also valid to say that late-time Hawking radiation is still entangled with the interior, exactly as Hawking (1976) predicted. So, Penington (2020) and Engelhardt and Folkestad (2022) recommend viewing black hole complementarity as the basis-dependent role of wormholes in evacuation.

Fuzzball Complementarity: Another type of evacuation solution that pursues a more radical break from general relativity is to argue that classical black holes are not even an effective description of spacetime. Mathur (2009) pioneers the proposal of stringy fuzzballs, an example of pseudo-black holes encapsulating non-generic extremal states in which the event horizon and interior never form in the first place. Even though extremal black holes don't evaporate, simple deviations from extremality generate 'Hawking-like' radiation. Now, there's no conflict between the global causal structure of spacetime and the eventual escape of radiation.

Nevertheless, this strategy requires novel physics in domains where general relativity was thought to work well, which is prima facie unattractive. Therefore, Mathur and Turton (2014) put forth a distinct notion of complementarity. A macroscopic inertial observer shouldn't be able to differentiate between the classical experience and quantum gravitational experience. The former entails free fall across the event horizon and into the black hole, followed by spaghettification due to tidal forces. The latter entails riding the collective vibration of strings constituting the surface system (which is all there is) before its own stringy constituents get tangled into the fuzzball.

[4]Bulk/boundary dualities technically relate distinct spacetimes differing by one dimension, whereas the black hole interior and Hawking radiation are in the same spacetime.

3.4.2 Interim Challenges and Responses: Horizons are Dramatic Cloning Devices

The elephant in the room for evacuation solutions is the implicit superluminal transmission of information/degrees of freedom across a macroscopic event horizon. To avoid outright postulating faster-than-light particles (like tachyons), evacuation solutions take advantage of three indirect evacuation mechanisms. First, the horizon can act as a "brick wall" and scatter infalling matter out to infinity. Nothing ever reaches the interior. Second, the horizon can act as a "bleaching agent". It purifies states that would otherwise be entangled across the horizon, and more generally, it transfers the wavefunction amplitudes of infalling matter to exterior degrees of freedom. Third, the horizon can act as a "cloning device". It records the signature of infalling matter, thereby allowing one copy of the system to pass freely inside and another copy of the system to escape (see Preskill 1992; Susskind and Lindesay 2004; Mann 2015). Each mechanism individually threatens the consistency of low-energy physics, and the proposals that I've laid out blend all three models, thus contributing to the misconception that evacuation solutions naively commit themselves to intolerable baggage.

Maudlin (2017), for one, is extremely unhappy with evacuation solutions, to the extent of scoffing at them as "solutions in search of a problem" (p. 17). All three mechanisms essentially force Σ_2 in Figure 3.1 to be a product state, which Maudlin (2017) accuses of being contrived to stymie entanglement across the black hole horizon and Unruh and Wald (2017) forcefully argue that quantum field theory does not sanction. Mann (2015) explains further,

> It seems straightforward that an astrophysical black hole can absorb an electron, so if this option holds then some new kind of physics – some kind of drama – must be present at the horizon to either prevent this from happening or to decouple the information in this state from its energy and angular momentum (whatever that means) (p. 68).

This "drama" undermines both general relativity and quantum field theory. On the one side, the brick wall and bleaching models are bemoaned to violate the equivalence principle. However, the equivalence principle is just a proxy to capture the condition that an event horizon is a global feature of spacetime, and classically, an inertial observer should not be able to tell it apart it from empty space. On the flip side, the cloning model is true to its name and violates the no-cloning theorem in a regime where it should be a decent approximation.[5] Critics of evacuation solutions mistakenly

[5]The no-cloning theorem exposes the incompatiblity between state duplication and linearity. For example, take a qubit state that's a superposition of a binary observable: $\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$. In order for an operator acting on the qubit state to produce a copy, we need the output to look like $\left[\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)\right]^2$, which involves cross terms. However, linearity demands that any operator acting on the entire state can also act on basis vectors independently, so we end up with $\left(\frac{1}{\sqrt{2}}|0\rangle\right)^2+\left(\frac{1}{\sqrt{2}}|1\rangle\right)^2$, which does not involve cross terms. Because cloning is nonlinear, it's also non-unitary.

view them as advocates of making Σ_2 a product state. Though Susskind et al. (1993) entertain the implications of that strategy, it's just their first stab at coming up with an evacuation solution and they too dismiss it as an "unreasonable violation" (p. 3747).

The biggest stumbling block for evacuation solutions is a foundational tenet of quantum field theory that was instrumental in Hawking's derivation and one that safehouse solutions hold on to: spacelike separated degrees of freedom are independent. This is a multifaceted statement commonly thought to underwrite causality/locality in relativistic spacetimes. The operational manifestation is that of the experimental independence of commuting observables, which precludes the intervention on a system in one location from affecting the statistical distribution of measurement outcomes of another system outside its lightcone. In a similar vein, cluster decomposition treats ingoing and outgoing degrees of freedom that are widely separated in space and time as non-interacting (see Raju 2022).[6]

The related Hilbert space manifestation warrants potential factorizability. Independent degrees of freedom of subsystems build up Hilbert subspaces with well-defined states. Entanglement between subsystems, which prevents actual factorization into a tensor product, doesn't threaten independence because what's relevant is the contributing dimensionality of the subspace to the overarching Hilbert space. It's enough for subspace states to be well-defined counterfactually in the absence of entanglement (see Earman 2015). The black hole interior is indeed spacelike separated from the exterior, so the presumption has naturally been that Σ_2 could hypothetically have been a product state between independent degrees of freedom, even though it may be physically unattractive to sever the entanglement. The way out then is to deny the independence of Σ_2's degrees of freedom, which Susskind et al. (1993) do on operational grounds.

> The assumption of a state... which simultaneously describes both the interior and exterior of a black hole seems suspiciously unphysical. Such a state can describe correlations which have no operational meaning, since an observer who passes behind the event horizon can never communicate the result of any experiment performed inside the black hole to an observer outside the black hole... [T]he state lying in the tensor product space $\mathcal{H}_{bh} \otimes \mathcal{H}_{out}$ can only be made use of by a 'superobserver' outside our universe (p. 3747).

The ramification of denying independence between $\mathcal{H}_{bh}$ and $\mathcal{H}_{out}$ is embracing redundancy alongside blatant nonlocality that's over and above entanglement, i.e., superluminal influence across spacelike separated regions.[7] In fact, by dispensing with

[6] Even though intervening on an entangled subsystem fixes the measurement outcome of its complement nonlocally, the statistical distribution over repeated iterations is unaffected.

[7] In the ER=EPR-inspired proposals, superluminal influence is traceable to wormholes complicating the domains of dependence of certain subregions. Since global hyperbolicity is supposed to guarantee that any subregion's domain of dependence adheres to local lightcone structure, the invocation of wormholes threatens global hyperbolicity. Yet notice that unlike in Hawking's derivation,

states like Σ_2, evacuation solutions have managed to score a two-for-one deal. They've identified the source of the impending nonlocality to evacuate degrees of freedom across a causal barrier. Furthermore, they've adopted some notion of complementarity to cash out the redundancy, which has the bonus of relegating the evacuation mechanisms' inconsistencies with low-energy physics to the realm of the "unphysical superobserver".

Early versions of black hole complementarity utilized non-commuting bases of subsets of interior and exterior degrees of freedom. The overarching Hilbert space is built up from either basis corresponding to the observables of a stationary or inertial frame of reference. Kiem et al. (1995) argued that to stay within low-energy regimes and avoid contradictions across reference frames, each basis must employ a frame-dependent UV-cutoff. It may seem like transforming across frames requires agreement about the presence of horizon hairs, but their respective UV-cutoffs ensure that stationary observers report a dramatic horizon whereas inertial observers don't (see Bokulich 2003; van Dongen and de Haro 2004).

Later versions of black hole complementarity, such as entanglement wedge reconstruction, import methods from AdS/CFT correspondence, but Penington (2020) acknowledges that AdS/CFT correspondence is put in by hand and not derived from first principles. Initially discovered by Maldacena (1999), AdS/CFT correspondence maps dynamics in an anti-de Sitter spacetime (which has a small, negative cosmological constant) to a conformal field theory in a lower-dimensional Minkowki spacetime. Witten (1998) also demonstrates that AdS/CFT correspondence satisfies the definition of theoretical equivalence between observables of the AdS bulk and those of the infinitely far away boundary, which is why it's also called a bulk/boundary duality. Raju (2022) motivates the relevance of bulk/boundary dualities for evaporating black holes: Gravity can't ever be screened off. Gravity thereby renders cluster decomposition moot, creating redundancies in arbitrarily distant degrees of freedom – akin to a cloning mechanism.

3.4.3 Bekenstein-Hawking Entropy: Duality between Interior and Exterior States

By design, evacuation solutions insist that by the time a black hole shrinks down to Planck mass, most of its degrees of freedom have been converted to radiation. This premise, along with the constraints introduced in Chapter 2, which state that maximal Boltzmann entropy is conserved and the von Neumann entropy of late-time Hawking radiation vanishes, leads to an alternative verdict about the interpretation of Bekenstein-Hawking entropy. It is indeed synonymous with overall black hole entropy, giving rise to unprecedented holographic information storage capacity. The key

non-global hyperbolicity in this context has nothing to do with failing to conserve degrees of freedom or entanglement across global states. In fact, all evacuation solutions are compatible with global unitarity, so as I foreshadowed in a footnote in Chapter 2, global hyperbolicity is indeed a sufficient but not necessary condition for its execution. Relaxing constraints on causal mellowness reveals a concrete example of when the latter can obtain without the former.

to isolating the departure from safehouse solutions lies in the information restoring function of the surface system.

The end goal of Susskind et al. (1993) is to show that for a stationary exterior observer, the process of infalling matter interacting with the black hole surface system and being emitted as radiation conserves maximal Boltzmann entropy. Therefore, in Hawking's derivation, all of the degrees of freedom of the post-evaporation state, Σ_3 in Figure 3.1, are independent and accounts for the true maximal Boltzmann entropy. The evolution from Σ_2 to Σ_3 only appears to cause a reduction in maximal Boltzmann entropy because Σ_2 overcounts independent degrees of freedom. In other words, the maximal Boltzmann entropy is much smaller than what Hawking originally calculated and what safehouse solutions purport because they've overlooked the redundancy between the black hole interior and exterior.

Exactly which subsets of degrees of freedom are redundant depends on the stage of black hole evaporation, particularly the von Neumann entropy of the radiation system, not just on the frame-dependent basis. During the early stages of evaporation, an exterior observer detects Hawking radiation that's maximally entangled with the black hole because the negative-energy partners haven't escaped yet, and in that stationary reference frame, a black hole amounts to a complex, quantum gravitational surface system whose interior is excised. An infalling observer, however, registers a surrounding vacuum and attests to the continuation of the black hole interior, as per the predictions of an inertial reference frame. So long as what Susskind (2013) refers to as the "proximity postulate" holds, in which interior degrees of freedom are reconstructed from near-horizon, exterior degrees of freedom, the story mirrors that of older versions of black hole complementarity.

During late stages of evaporation, however, Hawking radiation is no longer maximally entangled with the black hole interior/surface system because many of its negative-energy partners have escaped. Page (1993b) foresaw that in order to drive the von Neumann entropy of late-time radiation to zero, reflecting its status as a global pure state, entanglement would have to be transferred exclusively to exterior degrees of freedom. After all, it's perfectly acceptable for a pure state without any external correlations to still exhibit internal entanglement. Yet the validity of the inertial description of a drama-free horizon hangs on maximal entanglement between Hawking radiation and the black hole interior, and unfortunately, Mathur (2009) and Almheiri et al. (2013) prove that maximal entanglement is a zero-sum game.

In the spirit of exploiting redundancy to skirt violations of low-energy physics, Susskind (2013) anticipates that the proximity postulate must eventually give way to something akin to a distance postulate, such that interior degrees of freedom are reconstructed from faraway exterior degrees of freedom. By incorporating non-perturbative, low-energy corrections to Hawking's original calculation, Penington (2020), Almheiri et al. (2020), and Raju (2022) affirm Susskind's intuition and underwrite the distance postulate with insights from AdS/CFT correspondence. Almheiri et al. (2021) build upon these results to derive the famous Page curve (see Page 2013; Wallace 2018), which graphs the von Neumann entropy of Hawking radiation in Figure 3.3.

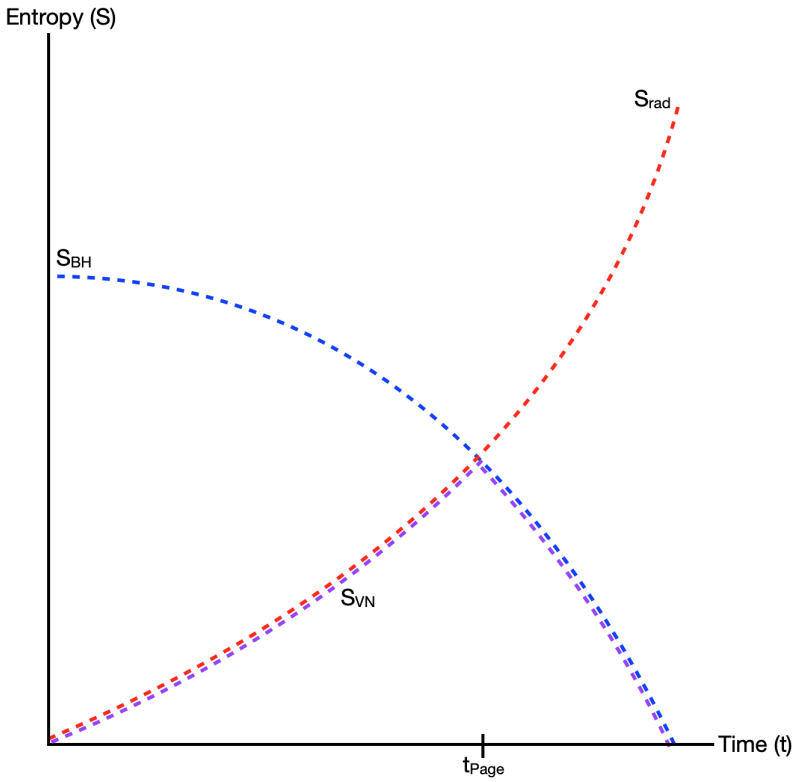

Figure 3.3: Evacuation Solutions: The Page Curve

The exterior radiation's thermodynamic entropy increases over time in discrete steps with the emission of positive-energy quanta, as shown by the red dashed curve. It's exactly thermal before the Page time and approximately thermal subsequently. The von Neumann entropy between the black hole and exterior radiation, as shown by the purple dashed "Page curve", increases in discrete steps with the production of Hawking pairs until the Page time, after which it decreases in discrete steps due to the transfer of trans-horizon entanglement to the exterior. The black hole's Bekenstein-Hawking entropy, as shown by the blue dashed curve, bounds von Neumann entropy from above. They are equal from the Page time onwards.

In contrast with the interpretation of safehouse solutions, S_{BH} (the blue dashed curve) is the Boltzmann entropy of the black hole interior, or equivalently, of the dual surface system via the proximity postulate. S_{VN}, the purple dashed curve, is the trans-horizon entanglement entropy that's always bounded from above by S_{BH}. Hence, S_{rad}, the red dashed curve, is only thermal (canonical) entropy up until the Page time when it coincides with S_{VN}. Subsequently, it morphs into microcanonical entropy.

The Page curve confirms that Bekenstein-Hawking entropy exhausts black hole degrees of freedom and is the logarithm of the dimensionality of its Hilbert space, even though the non-commutativity between interior and exterior observables prohibits localizing them (Page, 1995). The proximity postulate is in effect while the radiation's von Neumann entropy is increasing but the black hole's Bekenstein-Hawking entropy is still greater, insinuating that it has sufficient information storage capacity to house

the entanglement of monogamous Hawking pairs.

Then the Page time, t_{Page}, at which the radiation's von Neumann entropy matches the black hole's Bekenstein-Hawking entropy, marks the saturation point of the surface system's information storage capacity. Through numerical methods, Page (2013) estimates that the Page time occurs 53.81% into the total evaporation time when Bekenstein-Hawking entropy has decreased by 40.25%. Once the surface system can't house the entangled partners anymore, it returns them to the distant exterior, which takes over as the dual to the black hole interior. Therefore, the Page time is the turnaround point when the radiation's von Neumann entropy begins decreasing, and it's also the transition from the proximity postulate to the distance postulate, which executes bulk/boundary dualities.

Proponents of evacuation solutions have been galvanized from the get-go by entropy bounds. They distrust infinite degrees of freedom, perhaps because thermodynamic equilibrium becomes unattainable with unbounded entropy increase, which is their main objection against the infinite storage capacity of safehouse solutions (see Preskill 1992; Polchinski 2017). An entropy bound on a closed physical system is based on constraints such as size, energy, particle species, and interactions. Assuming that black holes are the most entropically dense objects in the universe, Bekenstein-Hawking entropy limits the number of degrees of freedom that can be packed into a region of spacetime (see Susskind 1995).

Page (1993a) calculated that for any finite, ergodic quantum statistical system partitioned into two initially unentangled subsystems, generic evolution exhibits a rise and fall in entanglement entropy along the Page curve. Ergodicity is determined with respect to the Haar measure, which tethers the time spent in a region of the Bloch hypersurface in Hilbert space to its volume. Haar-randomness guarantees that entanglement will grow for as long as possible before reversing, to rule out the special case that the initially unentangled subsystems stay pure the entire time. Drawing upon the example of Hawking pairs, Haar-randomness makes it extremely unlikely for entangled partners to escape consecutively. It follows then that the primary justification for evacuation solutions is an adaptation of what Susskind (1995) has coined the Holographic Principle (HP).

> **Holographic Principle (HP)**: Black holes are finite, ergodic, quantum statistical systems bounded above by their Bekenstein-Hawking entropy. Since Bekenstein-Hawking entropy is proportional to horizon area, it represents a holographic bound on entropy density with non-localizable degrees of freedom.

The metaphor of Bekenstein-Hawking entropy as a holographic bound on thermodynamic, statistical, and entanglement entropy is the evacuation solutions' claim to fame. The concept of holography was initially proposed by Bekenstein (1974) (loosely speaking), then further refined and popularized by Susskind (1995), and ultimately generalized and solidified by Bousso (2002). Bousso (2002) demonstrates that HP can be extended to a universal entropy bound in which the information storage cap on

any physical system enclosed in a region of spacetime is given not by the interior volume but by the area of a suitably-defined covariant surface. Therefore, the maximum amount of information needed to specify what is happening in the bulk of a system can be encoded on its boundary, much like a three-dimensional hologram is projected from a two-dimensional film. This analogy has been further sensationalized by the preliminary successes of AdS/CFT correspondence given the duality of observables between spacetimes differing by one dimension (Witten, 1998).

As with safehouse solutions, the nature of Bekenstein-Hawking degrees of freedom is going to depend on the theory, though physicists working on evacuation solutions tend to be heavily influenced by string theory. Strings are nonlocal objects whose spacelike separated ends have non-commuting observables. Low-energy commutators are strongly suppressed in flat spacetime but amplified when stretched in the vicinity of a black hole (see Lowe et al. 1995; Belot et al. 1999). 1996 was a formidable year for black hole statistical mechanics not just in loop quantum gravity but also in string theory when Strominger and Vafa (1996) enumerated the bulk microstates of extremal black holes in agreement with Bekenstein-Hawking entropy.

The takeaway of evacuation solutions is that the existence of the black hole interior is inextricably linked to the existence of the event horizon, whose eventual disappearance clocks the end of the entire black hole. That is why the interior and surface system aren't independent subsystems with additive entropies, as safehouse solutions suggest. Furthermore, bulk/boundary dualities entail that the distant exterior isn't an independent subsystem either. Bekenstein-Hawking entropy camouflages an intricate story of how the "same bit" admits of a multiplicity of physical realizations (Polchinski, 2017).

The only evacuation solution that I'm aware of which defies a strong holographic interpretation is fuzzball complementarity, and indeed, it shares several traits with safehouse solutions. By dispensing with the event horizon altogether, it saves itself the hassle of nonlocally evolving ingoing modes into outgoing Hawking radiation.[8] Furthermore, by blocking the emergence of classical spacetime behind the horizon (see Mathur 2009; Huggett and Matsubara 2021), it restricts Bekenstein-Hawking entropy to surface degrees of freedom. As far as the interpretation of Bekenstein-Hawking entropy goes, fuzzballs straddle the two camps, but conceptually, they align with the evacuation mentality in that final Hawking radiation is a global, and therefore, isolated system because there's no black hole interior to serve as a safehouse.

3.5 Advantages of the Safehouse/Evacuation Dichotomy: Guiding Principles

With a comprehensive synposis of the preeminent prospective solutions under our belt, the natural next step is to ascertain whether novel physics resolving phantom entan-

[8] However, Belot et al. (1999) assert that far from eliminating nonlocality, fuzzball complementarity insidiously smuggles it back in by preventing the teleological event horizon from forming.

glement pose novel problems for black hole evaporation. So, this is an apt juncture in the analysis to broaden the formulation of the black hole information loss paradox.

3.5.1 Selects Guiding Principles

The dichotomy between safehouse and evacuation solutions aids in optimizing the expanded formulation of the black hole information loss paradox because it exposes ambiguities in what had been assumed to be common vocabulary. Before we sit down and decide on the premises, there are several factors to keep in mind. What we should not do is rely on mathematical definitions and theorems of predecessor theories to frame new conflicts, such as the equivalence principle and global hyperbolicity in general relativity, as well as global unitary evolution in quantum theory. The more technical and outdated they are, the easier it is to dispute the premises as unfaithfully representing the diverse range of proposals. The form of the dynamical laws should also be left open to be inclusive of multiple research paradigms.

Rather, it's a feature, not a bug, to keep the premises flexible and open to interpretation, and since we're ultimately dealing with quantum gravity, Crowther (2018a) informs us that guiding principles are the best candidates. She distinguishes among several types of guiding principles, including 1) heuristics/philosophical assumptions/theoretical virtues for theory construction and development, 2) criteria for theory acceptance/success, and 3) means of non-empirical confirmation.

The first type of guiding principle usually deals with constraints internal to quantum gravity, such as the holographic principle. It's vital to the theory coming together and is the ladder that may or may not be kicked away in the end. The second type of guiding principle usually deals with constraints external to quantum gravity. Crowther (2018b) describes these as inter-theoretical guiding principles: 1) the generalized correspondence principle, which matches the predictions of distinct theories in overlapping domains, and 2) reduction, a specialized correspondence principle when one theory is relatively more fundamental than the other. Inter-theoretical guiding principles set thresholds for pursuit worthiness. The third type of guiding principle raises our credence in the theory after it has been provisionally accepted and often employs meta-inductive reasoning (see Dawid 2019).

As I will show, the terms that benefit from a looser reading in particular include 'semi-classical gravity', 'empirical adequacy', 'universality', and of course, 'Bekenstein-Hawking entropy'. The lack of consensus indicates that anyone engaging with these terms is responsible for filling in the details, and there are predictable patterns in how safehouse solutions go about it as opposed to evacuation solutions. Other critical concepts leave less room for negotiation, like 'effective theory', 'Planckian regimes', and 'inter-theoretical correspondence'. It would be quite difficult to equivocate on conditions that stipulate scale and relative fundamentality.

Nevertheless, given the precariousness of the status quo in quantum gravity, realistically what's going to happen is that one person supporting safehouse solutions will point to the desiderata and defend why they're jointly consistent for their chosen

proposal, while someone else supporting evacuation solutions will point to the same proposal and counter why it falls short, thereby contradicting said desiderata. But I also proffer recommendations to bridge the aisle so that these desiderata eventually whittle down the proposal space.

3.5.2 Deflates the Centrality of Unitarity

One of the best outcomes of the phantom entanglement framework is its independence from unitarity. Recall from Chapter 1 that I argued against the centrality of unitarity to the black hole information loss paradox. Nowhere in this analysis was it vital for the solutions to restore unitarity to solve phantom entanglement. That's because the contradiction revolved around a kinematic issue – the extent to which late-time radiation is entangled externally and its appropriate global state specification, which boils down to the cardinality of black hole degrees of freedom and its Hilbert space dimensionality. Therefore, it makes complete sense that the interpretation of Bekenstein-Hawking entropy, a kinematic quantity, drives the resolution. My recommended taxonomy of safehouse versus evacuation solutions places that revelation at the forefront.

Of the four information conservation principles that I laid out in Chapter 1, the only one that's truly at stake in the black hole information loss paradox is the condition of global physical statehood.

> **Global Physical Statehood (GPS)**: All global states are informationally complete.

Any potential solution rectifies the informationally incomplete status of externally entangled global states by ensuring that at whatever moment there are nonlocal correlations between the black hole and Hawking radiation, both subsystems have sufficiently large entropy to purify each other. Safehouse solutions decide not to restrict the black hole's purifying capacity just to Bekenstein-Hawking degrees of freedom, whereas evacuation solutions do.

Even though all of the proposals I've discussed are certainly compatible with other global information conservation principles, such as the conservation of degrees of freedom and deterministic dynamics (which jointly contribute to the constancy of coarse-graining), they're not necessary, strictly speaking. As long as information deficits about global states reveal epistemological, not ontological limitations, GPS is agnostic about the dimensional stability of the state space representation or the nature of laws. Nevertheless, varying degrees of freedom risks empirical adequacy, such as observed CPT invariance (see Page 1995), so it's probably safer to conserve them.

The urgency with which the mainstream narrative advocates for deterministic dynamics doesn't stand up to scrutiny when confronted with GPS, as I exposited in Chapter 1. Such flexibility precipitates a major advancement in the black hole information loss discourse because it opens up the debate to non-unitary quantum theories, which have overwhelmingly been left out. By extension, the focus on GPS also corrects the fallacy that embracing non-unitarity, particularly indeterministic dynamics,

addresses the black hole information loss paradox and the measurement problem in one fell swoop, as Okon and Sudarsky (2017) allege. The paradox of phantom entanglement reaffirms that black hole information loss is much deeper than individual metaphysical preferences about the ideal form of laws of nature.

3.5.3 Redefines Semi-classical Gravity

As opposed to unitarity in isolation, the paradox of phantom entanglement puts pressure on semi-classical gravity, even after its resolution. All of the proposals laid out draw inspiration from quantum gravity approaches, but in order to solidify their foundation for black hole evaporation, they need to redefine what it means to be a semi-classical framework, whose trustworthiness and temporary utility is up in the air (see Wüthrich 2021; Großardt 2022). An immediate benefit of grouping proposals as safehouse versus evacuation solutions is the clearcut divide in how they go about modifying Hawking's derivation and Penrose diagram to make black hole evaporation self-consistent.

Let's start with what safehouse and evacuation solutions must agree on for any revitalized semi-classical framework to deserve its designation. It's an effective theory in which spacetime is approximated as classical and the matter sector is described by quantum field theory, with mutual back-reaction (see Wald 1994; Großardt 2022). Conventionally, the coupling is encoded in the Semi-Classical Einstein Field Equation (see Equation 3.5), where the classical Einstein tensor, $G_{\mu\nu}$, satisfies the Einstein equations; however, the classical stress-energy tensor is replaced with the expectation value of a quantum energy-momentum operator defined on the matter fields, $\left\langle \hat{T}_{\mu\nu} \right\rangle_{\psi}$:

$$G_{\mu\nu} = \frac{8\pi G}{c^4} \left\langle \hat{T}_{\mu\nu} \right\rangle_{\psi} . \tag{3.5}$$

The domain of validity is sub-Planckian, and subleading perturbative corrections are meant to maintain the integrity of the theory (see Mathur 2009). I've compressed the minimal expectations of any iteration of semi-classical gravity in a criterion of theory acceptance/success.

> **Semi-classical Validity (SCV)**: Semi-classical gravity is an effective theory that couples aspects of general relativity and quantum field theory; it's valid at most up to perturbative corrections in sub-Planckian regimes.

I purposefully distinguish aspects of general relativity and quantum field theory from the theories themselves because we've already seen how simply smashing them together without much tweaking culminates in the paradox of phantom entanglement. Semi-classical gravity is supposed to be an independent effective theory with room to give up hallmark features of its predecessors. Therefore, choices must be made.

Below I've compiled comprehensive lists of classical and quantum ingredients that Hawking (1975, 1976) relied upon, which I'll identify as semi-classical gravity*, the

historically conventional formulation of semi-classical gravity (what Mattingly 2009 calls the "naive semiclassical Einstein theory") adapted to the (problematic) context of black hole evaporation. Something has to give because not everything on these lists must or can be assumed to hold concurrently.

Classical Ingredients of Semi-classical Gravity*:

1. **Scale**: Local radii of curvature larger than the Planck length (10^{-33} cm)
2. **Prerequisites for spacetime metric**: 1) Unquantized; 2) Continuous manifold outside of E; 3) external field
3. **Features of black holes**: 1) Global event horizon; 2) Singularity; 3) No-hair theorem
4. **Causality/Locality**: Speed-of-light barrier for causal processes
5. **Equivalence Principle**: Horizon crossing is locally indiscernible for inertial observers

Quantum Ingredients of Semi-classical Gravity*:

1. **Scale**: Planck energy of 10^9 J is upper bound
2. **Definition of stress-energy**: Expectation value of an energy-momentum operator defined on matter fields
3. **Prerequisites for vacuum state**: 1) Precludes further particle annihilation; 2) Uniqueness
4. **Causality/Locality**:
   (a) Vanishing (anti)commutators establishing experimental independence for spacelike separated observables
   (b) Potential factorizability of Hilbert space into black hole interior and exterior subspaces
   (c) Cluster decomposition treating spatiotemporally distant degrees of freedom as non-interacting
5. **Entanglement**: Field modes decomposed into non-separable linear combinations of positive and negative frequencies
6. **System-Field Coupling**: Stationary exterior reference frames decompose field modes discontinuously across the event horizon, which results in the presence of particles
   (a) Positive frequencies with respect to future infinity

(b) Negative frequencies with respect to the black hole interior

7. **Adiabatic Principle**: Inertial reference frames decompose field modes continuously across the event horizon, which suppresses the presence of particles

Given the length of these lists and the abundance of choices, it would seem that the semi-classical designation permits considerable latitude. What then has to give? Different proposals make different choices about what to surrender and what to retain. Many even add novel ingredients, especially those incorporating bulk/boundary dualities. So in the end, that's a trick question! The proliferation of proposals signals widespread disagreement on what has to give.

One might grumble that SCV is so malleable that it's essentially vacuous. We can't actually determine whether or not SCV features in a genuine contradiction. But the danger of pinning down the scope of semi-classical gravity is making it a vulnerable target for attack and unduly extending the lifetime of the black hole information loss paradox. There would be no hope for building consensus on the status of the paradox, and it would be a foregone opportunity to learn about dissenting background assumptions. Consequently, the most we should impose are the conditions for any legitimate semi-classical framework. Regardless of the particulars, we know that it's an effective, low-energy theory, and it becomes definitively unreliable at Planckian scales where high-energy quantum gravity kicks in.

Nevertheless, the objection is well-founded, revealing the need for another principle to weed out illegitimate frameworks masquerading as semi-classical contenders. Nothing trumps the epistemic virtue of empirical adequacy, and unlike the situation for quantum gravity, we have some empirical access to semi-classical gravity, with predictions confirmed in cosmology and astrophysics (see Wallace 2022). The plasticity of SCV should not be able to undermine observational data where semi-classical gravity* or predecessor theories have fared successfully.

> **Empirical Adequacy Condition (EAC)**: Empirically-confirmed predictions are recovered in their respective regimes.

Believe it or not, Susskind and Thorlacius (1994) vouch for EAC with thought experiments, another demonstration of their epistemic value à la El Skaf and Palacios (2022). They appreciate that the predictions of SCV will likely deviate from semi-classical gravity* even in sub-Planckian regimes, and they set up experimental situations to detect anomalies. They aver that if probing energies venture into Planckian regimes, then we've safely and swiftly switched from SCV to quantum gravity. Bokulich (2003) qualifies in response that "verification is not impossible, but is merely beyond the domain of applicability of our current physical theories" (p. 189). He also expresses discontent over this caveat.

> The fact that it would take Planck-scale energies to experimentally *verify* this low-energy description seems to be irrelevant. At the very least, we are owed an account of why considerations of the energies required should be a decisive factor in evaluating the [proposal] before us (p. 193).

Although I'm not condoning the use of operationalism as a strategy for metaphysical theorizing, I can offer insights into the utility behind Susskind and Thorlacius's thought experiments. Considering the mountain of evidence supporting the standard models of particle physics and cosmology (such as the cosmic microwave background radiation), any SCV framework bears the burden of demonstrating why we have not encountered violations in our low-energy laboratories or why we could not plausibly encounter violations in more exotic low-energy environments. The response is that if deviations between semi-classical gravity* and SCV demand Planckian energies to detect as well as a theory beyond the scope of SCV to properly account for, then the updated framework has offered the most attractive way out of phantom entanglement – we get to hold onto black hole evaporation as well as "effectively" keep our effective, low-energy quantum field theory.

Undoubtedly though, thought experiments only go so far. EAC gains the most legitimacy through actual experiments. Thébault (2019) advance a promising avenue to confirm semi-classical frameworks: analogue experimentation. For example, analogue experiments for Hawking radiation involve creating event horizons in other wave media, such as sound, to test the reflection of the relevant quanta, like phonons. Unfortunately, the strength of EAC vis-à-vis analogue experimentation is still not decisive.

Thébault (2019) optimistically proclaim that results like the spectra of reflected phonon modes establish inter-type uniformity between black holes and other macroscopic systems with horizons. Crowther et al. (2021), on the contrary, protest that analogue experimentation loses its credibility when we don't have empirical justification for inter-type uniformity in the first place. Even so, EAC is an improvement on SCV alone, and for the sake of this discussion, it's straightforward to grant SCV to all safehouse and evacuation solutions that model black hole evaporation as a dynamical process in (emergent) spacetime, which gets rid of a stalling pain point in consensus-building.

Now, it's inevitable that proposals adopting SCV in lieu of semi-classical gravity* to eliminate phantom entanglement will introduce novel physics. There are four scenarios in ascending order of urgency and insurgency in which adapting the original framework sustains the semi-classical designation. First, the novel physics perturbatively correct Hawking's calculation; they don't replace it. Second, the novel physics are siphoned off to Planckian regimes. Third, the novel physics come into play after evaporating black holes reach Planckian size and only affect the byproducts of evaporation, even if the impact is felt subsequently on larger, lower-energy scales. Fourth, if the novel physics do replace Hawking's calculation, 'Hawking-like' radiation is derived from scratch.

To start, safehouse solutions are committed to safeguarding Hawking's semi-classical framework to the extent that they can. Hsu (2007) avers,

> This scenario leads to a resolution of the paradox without non-locality or modifications of low energy physics (p. 67).

In that vein, safehouse solutions target Planck-scale physics around the final evapora-

tion event and the singularity for modifications. No amount of perturbative corrections will neutralize the singularity's annihilating effects, so safehouse solutions rely on the second and third strategies of having novel physics either constricted to the Planck scale or made manifest during late-stages of black hole evaporation to resurrect remnants.

Since the second and third strategies are mutually exclusive, not all safehouse solutions reject the same classical and quantum criteria. Massive remnants generally adopt the second strategy. Stable massive remnants reject continuous radiation for stationary observers after the Hawking temperature hits the upper, sub-Planckian energy bound. And decaying massive remnants reject the globality of the event horizon, which becomes notoriously fuzzy at the Planck radius.

Massless remnants, on the other hand, frequently adopt the third strategy. Causally-connected massless remnants reject the uniqueness of the vacuum as well as classical prerequisites for the spacetime metric. They adopt a discrete, granular structure at the fundamental, quantum gravitational level that's only continuous at the coarse-grained, semi-classical level. Causally-disconnected massless remnants, however, reject continuous spacetime metrics even at the coarse-grained level when black holes transform into baby universes pinching off from their parents. In both of these scenarios, the modifications play a critical role in explaining the phenomena only after an evaporating black hole has already entered Planckian domains.

Evacuation solutions, in contrast to safehouse solutions, are all cornered into executing more or less the same strategy to alter semi-classical gravity* and bypass phantom entanglement. They target their modifications to the near-horizon region to get trapped degrees of freedom out. The hope was to adopt the first strategy and purify thermal Hawking radiation with perturbative corrections, but that turns out to be a nonstarter (see Page 1995; Mathur 2009; Wallace 2020). The first strategy was doomed to fail anyway without singularity resolution, just like for safehouse solutions. The fact that both safehouse and evacuation solutions can't get off the ground without singularity resolution illuminates its indispensability as a guiding principle. More will be said about that in Section 3.5.6.

The second and third strategies also fail for evacuation solutions because novel physics kick in from the beginning, with the proximity postulate holding before the Page time and bulk/boundary dualities holding after the Page time. That's why evacuation solutions must resort to the fourth strategy and redo the calculation, more or less from scratch. For a long time in the discourse, the fourth strategy was perceived as giving up on semi-classical gravity wholesale because of the implications for macroscopic nonlocality and superluminal influence. According to Preskill (1992), "At the very least, the semiclassical picture of the causal structure must be very misleading" (p. 5).

This sentiment has changed with respect to newer iterations of semi-classical gravity employing nonperturbative corrections with gravitational path integral calculations and bulk/boundary dualities (see Penington 2020; Almheiri et al. 2021). The success of toy models adhering to SCV and taking advantage of AdS/CFT correspondence

has swayed many scholars that semi-classical gravity is a holographic theory plausibly negotiating with causality/locality, foreshadowing the same in quantum gravity (see Linnemann and Visser 2018). All evacuation solutions tied to the story of black hole complementarity end up rejecting similar classical and quantum criteria, including prerequisites for the spacetime metric, no-hair theorem, and causality/locality in their classical and quantum incarnations, to name a few.

3.5.4 Uncovers a Family of Nested Black Hole Paradoxes

As I asserted in Chapter 2, the predominant proposals situated in various quantum gravity approaches have unintentionally dealt with the paradox of phantom entanglement and bent the guiding principle of semi-classical validity (SCV) to their needs. The significance of this observation is that any problems arising within those proposals, such as the Page-time paradox, are derivative and do not occur at the level of information loss aptly captured as phantom entanglement. Consequently, arguments that deflate information loss in favor of other flavors of black hole paradoxes fall flat.

For instance, Wallace has catalyzed a paradigm shift in philosophy of physics by reframing black hole information loss as the Page-time paradox.

> I distinguish between two versions of the black hole information loss paradox. The first arises from apparent failure of unitarity on the spacetime of a completely evaporating black hole, which appears to be non-globally-hyperbolic; this is the most commonly discussed version of the paradox in the foundational and semipopular literature, and the case for calling it 'paradoxical' is less than compelling. But the second arises from a clash between a fully statistical-mechanical interpretation of black hole evaporation and the quantum-field-theoretic description used in derivations of the Hawking effect. This version of the paradox arises long before a black hole completely evaporates, seems to be the version that has played a central role in quantum gravity, and is genuinely paradoxical... The (mathematical) evidence against information loss advanced by physicists is much more naturally understood in terms of the second version of the paradox (Wallace, 2020, p. 209-10).

Contrary to the assertion that the first version of black hole information loss is "less than compelling", I spent a great deal of effort in Chapter 2 backing up its paradoxical status by revealing and elucidating the incoherence of phantom entanglement. Now the question is whether the distinct Page-time paradox can persuasively supersede it, as Wallace (2020) contends.

Although I plan to analyze the Page-time paradox in greater depth in future work, here's a summary of its implications. Almheiri et al. (2013) expose the failure of the proximity postulate in black hole complementarity after the Page time. The purification of Hawking radiation requires entanglement between near-horizon degrees of freedom and distant degrees of freedom, but due to the monogamy of entanglement,

near-horizon degrees of freedom can no longer be entangled with interior degrees of freedom. This scenario disrupts the near-horizon vacuum and results in highly energetic modes just outside the black hole. As a result, the loss of trans-horizon entanglement releases divergent energy and generates a "firewall".

Yet the attempt to duplicate entanglement after the Page time to tame fiery drama is tantamount to cloning. Black hole complementarity, a clever ruse for cloning, can't come to the rescue and siphon off deviations from low-energy physics to the supposedly unphysical superobserver or quantum gravitational sector. That's because violations of the no-cloning theorem are detectable along the worldline of a single, low-energy observer. So without the resources of stealthy cloning, complementarity actually engenders a contradiction, in which infalling matter both is and isn't incinerated upon interacting with the infernal surface system, culminating in the Page-time paradox (see Polchinski 2017; Wallace 2020).

Bousso (2013) is one of the few vocal advocates of firewalls as a legitimate evacuation solution in and of itself. He alerts us to naively expecting the adiabatic vacuum at the horizon throughout black hole evaporation. After the Page-time, perhaps a new type of naked singularity emerges, or better yet, the black hole interior fails to emerge (see Polchinski 2017; Huggett and Matsubara 2021).

The main objection against firewalls is that they undermine the equivalence principle at the event horizon. Crossing the edge of a black hole is supposed to be indiscernible from free fall in empty space. But for a proponent of firewalls, the rebuttal to this objection is that the equivalence principle is no longer violated where it's not applicable, i.e., where there's no classical spacetime.

However, those who do not wish to retreat so desperately by having high-energy quantum gravity severely encroach upon the domain of semi-classical gravity and general relativity are motivated to resolve the Page-time paradox without resorting to firewalls. The substitution of the proximity postulate with the distance postulate after the Page time, a move supported by bulk/boundary dualities, is meant to defuse the firewall argument (see Maldacena and Susskind 2013; Penington 2020; Raju 2022).

However, the tenability of the radical nonlocality inherent in the distance postulate has stalled consensus, leaving the status of the Page-time paradox unsettled (Polchinski, 2017). Furthermore, Susskind (2008) has sensationalized the conflict as a paradigmatic clash between unitary evolution from quantum theory and the equivalence principle from general relativity. This spin on the Page-time paradox has since been rehearsed by physicists such as Bousso (2013) and philosophers such as Wüthrich (2021), thereby obscuring the presumed interpretation of Bekenstein-Hawking entropy.

To clean up the discourse, what Wallace (2020) does absolutely right is casting the Page-time paradox as a conflict between black hole statistical mechanics and quantum field theory on curved spacetime. Yet what he leaves out is that the conflict doesn't encompass the entire field of black hole statistical mechanics. Since the Page curve of Figure 3.3 pertains specifically to the holographic execution, the Page-time paradox bears primarily on evacuation solutions.

Most safehouse solutions are immune from the firewall argument, with the excep-

tion of decaying remnants whose enormous but finite entropy is eventually outpaced by the growth of trans-horizon entanglement. These remnants may confront fiery drama well after the Page time when the interior degrees of freedom purifying Hawking radiation are exhausted (Chen et al., 2015), though ideally this happens when the evaporating black hole has already shrunk down to Planck mass so that the paradox loses its bite regarding the premature upending of SCV.

Nevertheless, such remnants of bounded internal entropy seem to be ruled out based on criticisms explored in Section 3.3.2. For that reason, the Page-time paradox can't possibly supplant the paradox of phantom entanglement, which is ultimately the foundational paradox about the interpretation of Bekenstein-Hawking entropy that has played a central role in the development and proliferation of proposals spanning various quantum gravity approaches.

Just because the Page-time paradox arises when a black hole has evaporated about halfway through and is still macroscopic doesn't mean that it precedes adjudication on post-evaporation information loss. Only after phantom entanglement has been dealt with by evacuating interior degrees of freedom to purify late-time Hawking radiation, thus pinning down the holographic interpretation of Bekenstein-Hawking entropy, does the Page-time paradox retroactively rear the horns of a trilemma: the "purity of the Hawking radiation, absence of infalling drama, and semi-classical behavior outside the horizon" (Almheiri et al., 2013).

Although a formulation of the Page-time paradox has been the topic of fierce debate for evacuation solutions, I have not encountered any discussion in the literature about how safehouse solutions deserve their own version of the Page-time paradox. Safehouse solutions aver that Hawking radiation is exactly thermal for the duration of the black hole evaporation conjecture (BHEC), at least until the evaporating black hole reaches Planck mass, thereby running into a sticky situation with respect to the foundations of statistical mechanics.

Thermal systems belong to statistical ensembles incorporating the Boltzmann energy distribution at fixed temperature. Classically that's the canonical ensemble, with a quantum mechanical analog involving a distribution over energy eigenstates. In either context, the canonical ensemble is appropriate for finite systems only when they're coupled to a relatively vast external reservoir serving as a heat bath to ensure rapid equilibration.

Even so, let's stretch the canonical ensemble's applicability to a heat bath that's at least comparable in size. Because safehouse solutions attribute thermodynamic black hole entropy to just the surface system carrying Bekenstein-Hawking degrees of freedom, we can infer that the surface system serves as an appropriate heat bath for Hawking radiation prior to the Page time so long as $S_{BH} > S_{rad}$ (see Figure 3.2).

In fact, the comparison between evaporating black holes and black bodies holds pretty tightly thus far due to the Causality Argument (CA). The surface system would count as a genuine black body because it's emitting thermal radiation that propagates in its future lightcone. It's actually warranted for a black body to serve as a heat bath since it absorbs and emits radiation at energies that are conducive to maintaining

thermal equilibrium.

After the Page time, however, when $S_B < S_{rad}$, the black hole surface system defies the traditional notion of a heat bath. Once an evaporating black body becomes smaller than the radiation subsystem, its ability to maintain true thermal equilibrium with the radiation is compromised. It' not able to mitigate energy fluxes in the absorption and emission of radiation to stabilize the temperature as efficiently as before. This can lead to temperature fluctuations and deviations from the idealized behavior of a heat bath. Therefore, the assumptions of the canonical ensemble cease to be valid and the black hole surface system should stop radiating like a perfect black body (see Hossenfelder 2004; Casadio and Harms 2011).

However, the continued thermality of Hawking radiation suggests it inexplicably continues to do so. To be honest though, even if we bite the bullet and compel the black hole surface system to produce thermal radiation despite the physical conditions for that idealization falling through, we'd be committing a bigger blunder and compromising the derivation of Bekenstein-Hawking entropy from the First Law. The First Law holds between two subsystems that experience energy transfers and then equilibrate at fixed temperature. Because the black hole surface system doesn't fully equilibrate with Hawking radiation after the Page time, the First Law becomes invalid, along with the derivation of Bekenstein-Hawking entropy from Hawking temperature.

Therefore, the unfettered entanglement between Hawking radiation and the deep interior does little to ameliorate the obstacle of a finite thermodynamic black hole subsystem. It conflicts with the definition of thermality while the black hole is still macroscopic and within the purview of semi-classical gravity, pitting black hole statistical mechanics against effective field theory approximations for safehouse solutions as well. Consequently, the dichotomy of causal entropy versus holographic entropy sets up a nested structure of black hole paradoxes, positioning the paradox of phantom entanglement as the bottom tier and the Page-time paradoxes a tier above.

3.5.5 Whittles Down the Proposal Space

The Page-time paradoxes illuminate that despite the plethora of prototypical solutions aiming to resolve the paradox of phantom entanglement, it's too soon to declare victory against black hole information loss. To start, numerous infelicities still need to be ironed out. Additionally, the burgeoning proposal space hampers progress because it signals widespread disagreement with the escalating polarization of philosophical commitments and a lack of consensus-building.

Nevertheless, the polarization of the interpretation of Bekenstein-Hawking entropy has a silver lining. If we have principled reasons to slash either safehouse or evacuation solutions as a category, we could swiftly and drastically curb the proposal space, thereby providing timely and much needed momentum to developing a functional theory of quantum gravity. Wallace (2020) figures just as much when he equates black hole statistical mechanics with evacuation solutions, which would explain why he elevated the conventional Page-time paradox as the premier black hole paradox

worthy of pursuit.

> Remnants, or thunderbolts, or baby universes, no matter how helpful they may be in preserving unitarity, do nothing to preserve the statistical interpretation of black hole entropy or any account of black hole thermodynamics as arising from statistical mechanics in the ordinary way, and so have no role in resolving this version of the information-loss paradox (Wallace, 2020, p. 222).

Wallace essentially sidelines safehouse solutions for allegedly compromising the statistical interpretation of Bekenstein-Hawking entropy and falling short in explaining black hole evaporation thermodynamically. The good news is that he's handed us exactly the principled reason we need on a platter to eliminate a whole class of proposals in one fell swoop. That principled reason is the extension of thermodynamics/statistical mechanics in their familiar, terrestrial applications to their unfamiliar, extraterrestrial applications involving black holes.

Crowther (2018a) observes how recovering Bekenstein-Hawking entropy through microstate enumeration is already a guiding principle for quantum gravity. She classifies the recovery of Bekenstein-Hawking entropy as a guiding principle functioning in all three of its capacities: 1) heuristic/philosophical assumption/theoretical virtue for theory construction and development, 2) criterion for theory acceptance/success, and 3) means of non-empirical confirmation. You may be wondering (as I sure was) how that doesn't end up being circular. If a guiding principle is used in theory construction and development, it doesn't make sense to turn around and say that it independently serves as a criterion for acceptance/success because it was built in from the get-go.[9]

But then I realized that the role of Bekenstein-Hawking entropy is multifaceted and linked with different guiding capacities. It's most notably an external constraint from semi-classical gravity because it embodies a mathematical principle. Bekenstein-Hawking entropy is a quarter of the horizon area to leading order, so deriving the proportionality factor is a litmus test for pursuit worthiness. This is why most scholars tend to view the recovery of Bekenstein-Hawking entropy as a criterion of theory acceptance/success, like Wüthrich (2019) and Huggett and Matsubara (2021).

However, in order for the recovery be more than a coincidence or a massaging of the formalism, the underlying model must convince us that black holes are truly thermodynamic systems with statistical mechanical underpinnings. Wallace's complaint about safehouse solutions insinuates that the role of recovering Bekenstein-Hawking entropy in raising our credence rests on the model's ability to deliver on the inter-theoretical guiding principles of correspondence and reduction. Thus far, we've outlined the end goals for Bekenstein-Hawking entropy that are summed up in what I refer to as the Universality Argument (UA).

[9]My thanks goes out to Lorenzo Cocco for helpful discussion about using guiding principles in a circular manner.

> **Universality Argument (UA)**: Black holes consist of Planckian degrees of freedom whose aggregate, statistical behavior recovers Bekenstein-Hawking entropy, which in turn reproduces universal thermodynamic phenomenology.

I realize that 'universality' has a technical connotation in statistical mechanics, in which renormalization group methods delimit a 'universality class' based on attractor dynamics in a reduced, coarse-grained phase space. I'm not necessarily using the term in that way, though it would be aspirational to lump together black holes and burning coal as such (where black holes are actually simpler systems than burning coal according to Raju 2022).

Batterman (2000) contends that the philosophical concept of multiple realizability is an instance of universality. Multiple realizability captures diverse systems' commonalities, which is explained by the stability of behaviors at the emergent level under perturbations at the fundamental level. This phenomenological stability is crucial to the conceptualization of universality and is ascertained through the successful application of "minimal models" to diverse systems, thereby allowing their streamlined representation (Batterman, 2019).

To that end, UA codifies the existence of commonalities between black holes and a wide range of physical systems based on minimal models of thermodynamics and statistical mechanics, such as the Page curve. We've already explored a few examples of universal behaviors when analyzing the position of Camp 2 in Section 3.2.1, such as the prerequisites of mediating thermal contact, equilibrating and settling into equilibrium, implementing a Carnot cycle, radiating the Planck spectrum as a black body, etc. UA also demands a reductive link between black hole thermodynamics and black hole statistical mechanics, although the exact details of that link hang on the identification of physically salient multiply realized phenomena among self-gravitating and other types of thermodynamic systems.

Juxtaposing the Causality Argument (CA) with the Holographic Principle (HP), it's obvious that there's a huge disparity between safehouse and evacuation solutions in picking out the relevant multiply realized phenomena. CA prioritizes the ability of black holes to mediate thermal contact locally and in a causally well-behaved manner, whereas HP protects the statistical definition of thermality and is under no illusion that a black hole can be modeled faithfully as a black body after the Page time.

Due to the nuances in discerning physical salience, I believe that Wallace (2020) and several physicists before him (see also t'Hooft 1996; Polchinski 2017) jumped the gun in dismissing safehouse solutions. The unambiguous distinction between terrestrial systems and black holes is that the latter are supposed to possess causal barriers, so cashing out out the requisite correspondence and reduction relations "in the ordinary way" is where the philosophical tension resides – hence, the ongoing heated debate over the correct interpretation of Bekenstein-Hawking entropy.

That's why the aspect of Bekenstein-Hawking entropy that's a guiding principle contributing to theory construction and development for black hole physics in quantum

gravity is the provisional commitment to CA or HP. Even though safehouse solutions tend to be nestled within loop quantum gravity and evacuation solutions within string theory, both approaches are hospitable to both camps.

For example, the Immirzi parameter is a free parameter in loop quantum gravity and can be fixed to reproduce holographic Bekenstein-Hawking entropy (see Gambini and Pullin 2008). On the other hand, AdS/CFT correspondence may not be a perfect duality with one-to-one mappings between bulk and boundary stringy observables; it's compatible with additional interior degrees of freedom that do not influence the exterior, and consequently, cannot be reconstructed from degrees of freedom elsewhere (see Marolf 2009; Hubeny and Rangamani 2012).

Therefore, the interpretation of Bekenstein-Hawking entropy is a hypothesis to be tested against specific criteria about the relevant multiply realized phenomena of UA – though laying out criteria such that both camps agree on a common set of standards is a nontrivial task. As I hinted at earlier, two preliminary criteria that pit safehouse solutions and evacuation solutions against each other are the capability of black holes to mediate thermal contact and the definition of thermality.

Proponents of safehouse solutions aver that complementarity-based proposals subvert standard accounts of thermodynamic equilibration. The frame-dependence of the allegedly thermalizing surface system in the membrane paradigm calls into question its dynamical reality, which Susskind and Thorlacius (1994) and Wallace (2018) concede yet deflate through operational arguments.

Curiel (2023a) also highlights that the equilibrating systems in question must be appropriately configured for mutual coupling. Complementarity excludes coupling between infalling observers and the infernal stretched horizon in the inertial description. The dual black hole interior available in the inertial description isn't an apt substitute either because it's never in thermodynamic equilibrium (see Sorkin 1997; Sorkin 2011). Therefore, it seems to me that evacuation solutions do indeed struggle to recover multiply realized thermodynamic phenomena.

On the other hand, proponents of evacuation solutions denounce safehouse solutions for their aberrant statistical foundations, though most of their concerns are for late-stage remnants. However, let me cursorily outline other concerns that I've not encountered in the literature. Back in Chapter 1, I briefly recounted how thermal systems in quantum field theory are typically entangled with their heat baths. This synergy demonstrably fails for safehouse solutions after the Page time as depicted in Figure 3.2, but it also fails beforehand.

CA typically attributes Bekenstein-Hawking entropy to trans-horizon entanglement over short distances, so those degrees of freedom were never available from the get-go to entangle with Hawking radiation. Therefore, the surface system can't be directly responsible for the sustenance of Hawking pairs even though it's somehow crucial to the radiation's thermodynamic behavior. It's then befuddling as to why Hawking radiation is in thermal equilibrium with one black hole subsystem – the surface system, but entangled with another – the deep interior. Consequently, it appears to me that safehouse solutions also falter in the reductive link between thermodynamic

phenomenology and Bekenstein-Hawking degrees of freedom.[10]

While the rivalry between safehouse and evacuation solutions has been given plenty of airtime, there's a shortage of arguments etched in print justifying one interpretation of Bekenstein-Hawking entropy over the other, save for the Socratic-style conversation of Jacobson et al. (2005) and more recently, the focused investigation of Engelhardt and Wall (2017) between CA and HP for recovering their respective entropy curves in AdS/CFT correspondence. Further work needs to be done to make either camp more convincing vis-à-vis UA in a systematic comparative analysis.[11]

UA also warrants a separate guiding principle, one that institutes generalized correspondence (see Crowther 2018b). UA invokes the high-energy physics of Planckian degrees of freedom, in contrast to SCV, which is limited to sub-Planckian, low-energy regimes. It's crucial for the sake of breadth and inter-theoretical consistency that the physics transition smoothly from whatever quantum gravity approach is underwriting UA to the chosen formalism of semi-classical gravity, and moreover, from semi-classical gravity to the established and empirically confirmed theories of quantum field theory and general relativity.

> **Generalized Correspondence Principle (GCP)**: Effective theories are recovered in the appropriate limits.

Nevertheless, satisfying GCP is an enormously tall order. The slight inconvenience is that we don't have a final theory of quantum gravity. None of the extant proposals stand out as advancing our understanding of black holes as composite quantum gravitational systems, not least because the results for statistically deriving Bekenstein-Hawking entropy are niche in both string theory and loop quantum gravity. Furthermore, we don't have a handle on the borders of our current theories, so who's to say, for example, that firewalls are paradoxical and not the most feasible candidate, as does Bousso (2013).[12] Nor can we fully anticipate complications for familiar regimes coming from novel quantum gravitational physics, such as the rich internal structure of remnants or the pervasive nonlocality of complementarity.

Without knowing anything more about the ontology of Bekenstein-Hawking entropy, the most sensible place to search for keys in the dark is – shocker – under the lamppost, which is indeed the Universality Argument (UA). Based on our current knowledge of thermodynamics and statistical mechanics, minimal models aid us in homogenizing diverse representations of black holes and catching threats to the end goal of multiply realizing physically salient behaviors, in spite of our nascent comprehension of the fundamental structure.

[10] I'd like to thank David Wallace for alerting me to the disassociation between thermal entropy and entanglement entropy for safehouse solutions.

[11] I'm very grateful to David Wallace for partaking in numerous conversations about UA for both evacuation and safehouse solutions.

[12] I'd like to thank Keizo Matsubara for discussion about the nuances of determining the domain of a theory. In the case of black holes, one may counter that without empirical access to the event horizon, it's fair game to impose a classical or semi-classical cut-off there and extend the domain of high-energy quantum gravity, such as with firewalls.

3.5.6 Strengthens the Case for Singularity Resolution

The guiding principle that brings together proponents of both safehouse and evacuation solutions in solidarity is singularity resolution. Now that may not appear to be an earth-shattering revelation because the discourse around singularity resolution has flourished independently of the black hole information loss debate, where Crowther and de Haro (2022) conclude that it's a somewhat inflated desideratum for quantum gravity. However, the paradox of phantom entanglement sheds new light on singularity resolution and raises its priority.

Crowther and de Haro (2022) differentiate between motivations for singularity resolution that are internal and external to general relativity. The main internal motivation is to fix incompleteness, given that general relativity does not produce unique solutions for singular spacetimes and the dominion of the Einstein Field Equation does not extend to the realm of singularities beyond the manifold. The main external motivations are that quantization has a track record of curing infinities, where quantum gravity intimates a minimal, Planck length, and corrections from semi-classical gravity already cast doubt on the accuracy of general relativity in pretty high-curvature regions around singularities. The expectation for quantum gravity is that singularity resolution invites mathematical consistency and signals novel physics.

Nonetheless, Crowther and de Haro are not quite persuaded that singularity resolution is such a strong a guiding principle for quantum gravity; they allege that external motivations are "more risky" than internal ones because they "stem from untested combinations of assumptions and heuristic arguments" (p. 245). And in agreement with Earman (1996), they find the main internal motivation, which carries the burden of persuasion, to be left wanting.

Earman (1996) proclaims that general relativity is incomplete only insofar as determinism is the hallmark of classical theories, which it doesn't have to be. But more to the point, how can a theory be incomplete in domains beyond its reach? General relativity is a theory of spacetime, and singularities fall outside the jurisdiction of spacetime.

Although I find the conflation between incompleteness and indeterminism to be a sleight of hand, for the sake of argument, let's accept that the case for singularity resolution thus far is flimsy. The paradox of phantom entanglement, however, revitalizes and injects force into the case for singularity resolution. It pivots the anchor for internal and external motivations from general relativity to semi-classical gravity*, the precursor to SCV.

The culprit of the contradiction in semi-classical gravity*, where late-time Hawking radiation is an entangled global system, is the black hole singularity. Phantom entanglement is a direct ramification of the singularity's annihilating effects. Therefore, singularity resolution is a necessary condition for black hole information conservation, as Hossenfelder and Smolin (2010) assert.[13]

[13]I know of only one potential counterexample. Maudlin (2017) follows Wald's lead by foliating a singular evaporation spacetime with disconnected Cauchy surfaces to conserve information in virtue

For safehouse solutions, eliminating the singularity extends the time available for black hole interiors to store degrees of freedom. For evacuation solutions, it clears the path to AdS/CFT correspondence: If the boundary is singularity-free, then so is the bulk. Seeing as there's no stronger internal motivation for singularity resolution than the dissolution of a paradox, both safehouse and evacuation solutions make sure that their elaboration of SCV is singularity-free.

Of course, black hole evaporation itself stems from untested assumptions and heuristic arguments, so why should we prioritize the paradox of phantom entanglement to justify singularity resolution? To answer that, let me first clarify how Crowther and de Haro (2022) are implicitly incorporating the guiding principle of regime-dependent empirical adequacy (EAC) into their argument.

Earman (1996) notes that singularities are unobservable features of spacetime, which I interpret as conveying that both internal and external motivations for singularity resolution are somewhat orthogonal to the empirical content of the theory under scrutiny. EAC cannot deliver a verdict between the presence and absence of singularities that would directly sanction singularity resolution.

The more powerful justification is discerning whether the purview of EAC for the observable entities of any given theory is sensitive to singularity resolution. According to Earman, Crowther, and de Haro, since general relativity is empirically adequate and self-consistent with singularities, internal motivations for singularity resolution are tenuous and EAC unaffected. General relativity's domain could plausibly be all of spacetime, right up to any precarious, singular edge.

However, the empirical adequacy of more fundamental theories whose predictions deviate from general relativity in high-curvature regions prior to singular behavior could prompt credible external motivations for singularity resolution. Put differently, if EAC were well-founded for semi-classical or quantum gravity, the borders of general relativity would have to be scaled back. We may already be there.

As I brought up previously, EAC is indeed reasonably supported for semi-classical phenomena in astrophysics, which on its own raises the credibility of external motivations for singularity resolution with respect to general relativity. But since we're pivoting the anchor to semi-classical gravity*, whose framework overlaps with the formalism checked against astrophysical observations vis-à-vis SEFE, let's re-evaluate how internal and external motivations for singularity resolution fare.

Semi-classical gravity* takes the arguments for tolerating singularities to heart and treats its classical domain as all of spacetime. Unlike the situation for general relativity, however, the presence of singularities in semi-classical gravity* prompts

of fully restoring global hyperbolicity. The goal is to construct a safehouse solution by demonstrating that interior modes, though they are eventually annihilated by the singularity, are not annihilated in the past lightcone of the final evaporation event. This apparent lack of causal separation frees the interior modes to purify exterior modes arbitrarily far into the future. Manchak and Weatherall (2018) have shown that this proposal is not viable for its self-professed aims since it doesn't actually alleviate the failure of global hyperbolicity. Whether this proposal nonetheless resolves phantom entanglement remains to be seen.

the paradox of phantom entanglement, rendering the theory inconsistent. Because of logical explosion, in which a contradiction implies anything and everything, the empirical success of semi-classical gravity* beyond black hole physics can no longer be credited to theory-specific predictions. The limited range of EAC for semi-classical gravity* has now been squashed to zero and any semblance of empirical adequacy completely nullified. The predominant internal motivation for singularity resolution in semi-classical gravity* – the protection of EAC – is thus more robust than that for general relativity.

On the other hand, Crowther and de Haro (2022) are justified in questioning external motivations from quantum gravity that still don't have recourse to EAC to appease doubts about risk. In spite of those doubts, deflating one of those external motivations that hasn't gotten proper attention in the literature would incur a significantly high cost because of its indispensability as a guiding principle. That guiding principle is none other than the universality argument for black hole thermodynamics and statistical mechanics (UA). If phantom entanglement is not taken care of via singularity resolution, then black holes can't evaporate in an appropriately semi-classical setting, i.e., a non-paradoxical successor to semi-classical gravity*.

Blocking the legitimacy of SCV and BHEC essentially undercuts UA, ushering in the downfall of black hole thermodynamics and statistical mechanics. SCV grounds black hole thermodynamics because the laws of thermodynamics always involve spatiotemporal phenomena and these particular phenomena involve low-energy quantum fields. BHEC also grounds black hole thermodynamics seeing as the latter is predicated on Hawking radiation possessing a physical temperature determined by the extent of energy flux across the horizon. It then goes without saying that ousting black hole thermodynamics thwarts reduction to black hole statistical mechanics in quantum gravity. Consequently, when SCV and BHEC crumble, UA crumbles in turn.

So without singularity resolution as a guiding principle, whatever theory construction and development hinging on UA that has advanced various quantum gravity approaches would be erased. For the leading contenders especially – string theory and loop quantum gravity – ample progress resulting from the execution of UA has reverberated outside the study of black holes, from the holographic universe conjecture in string theory to the Big Bounce model in loop quantum cosmology (see Bousso 2002; Ashtekar and Singh 2011). Ultimately, the strongest external motivation for singularity resolution in semi-classical gravity* is that of a subsidiary guiding principle in service of UA.

3.5.7 Strengthens the Case for Unification

Another downstream effect of UA is solidifying the guiding principle of unification at the fundamental level, again with predictable patterns between safehouse and evacuation solutions. This consequence is timely because the last 15 years or so have seen a backlash against the gridlock in quantum gravity, where scholars have contemplated whether the so-called virtue of unifying the fundamental ontology has hampered

progress. Hossenfelder (2018) nods vigorously in the affirmative and cautions general audiences in her book, *Lost in Math: How Beauty Leads Physics Astray*, about theoretical physics's obsession with the mathematical beauty associated with unification.

Wüthrich (2005) explains how unification at the fundamental level assumes that all degrees of freedom are quantum, in which the metric is either explicitly quantized or emerges from non-spatiotemporal quantum degrees of freedom. He also remarks that "unification for the sake of unification... does not sway the skeptic" (p. 778). In a similar vein, Mattingly (2005) rebuffs the payoffs of unification identified in philosophy of science and metaphysics, such as ontological parsimony, universal laws, etc. Mattingly (2009) goes onto argue that semi-classical gravity should be pushed to its brink, and while he agrees with SCV in terms of the latitude in its designation, he disagrees with the Planck-scale cutoff.

I concur that bottom-up reasoning for a unified theory of quantum gravity is not decisive unless tight parallels can be made with prior cases of successful unification in a meta-inductive fashion, à la Dawid (2019). UA, however, offers top-down reasoning for unification in quantum gravity. As I brought up in Section 3.2.2, the multiple realizability of thermodynamic behaviors is a manifestation of unification at the coarse-grained level with similar schemes for reduction to statistical mechanics. Large systems have similar macroscopic degrees of freedom because their microscopic degrees of freedom are interacting in ways made probable by attractor dynamics, or to use the colorful terminology of Curiel (2023a), they're "wiggling and jiggling". Minimal models in thermodynamics and statistical mechanics have widespread applicability across theories, gesturing towards hefty meta-inductive support for practically all systems other than black holes, ranging from engines to living organisms to stars.

So, in order for UA to go through, the explanation for Hawking radiation has to be the "wiggling and jiggling" of black holes' microscopic constituents. Yet it's not enough for these microscopic constituents to be quantized matter fields bounded by classical spatiotemporal regions, as Jacobson conjectures in Jacobson et al. (2005), because then black holes wouldn't evaporate. Curiel (2023a) raises the powerful insight that if there's truly a transfer of energy between black hole mass and Hawking radiation, then there must be black hole degrees of freedom capable of interacting with radiation degrees of freedom. And since the transfer of energy hangs on the presence of trans-horizon entanglement, both subsystems must be inherently quantum. Therefore, unification at the coarse-grained level implies unification at the fine-grained level.

Safehouse solutions, which find themselves at home in loop quantum gravity, implement the unification project by discretizing spacetime. Black holes are made of quantized metrical degrees of freedom, i.e., atoms of geometry, that interact and entangle with quantized matter degrees of freedom. But evacuation solutions, for whom string theory is the comfort zone, take the unification project further. UA doesn't just compel Bekenstein-Hawking degrees of freedom to be inherently quantum. The transmutation of spacetime into matter also resonates with unifying the ontology. Fundamentally, all degrees of freedom are non-spatiotemporal, and the distinction between matter and spatiotemporal degrees of freedom is merely emergent.

3.6 Conclusion: Transcending Black Hole Paradoxes

Upon concluding the analysis of guiding principles for prospective information conservation proposals, we're now primed to put forth the minimal premises that must be jointly respected to evade a paradox. I'm packaging them together as the "Bare Desiderata to Transcend Black Hole Paradoxes", which are nevertheless challenging to make consistent.

Bare Desiderata to Transcend Black Hole Paradoxes:

1. **Black Hole Evaporation Conjecture (BHEC)**: Black holes evaporate at least until reaching Planck mass.

2. **Global Physical Statehood (GPS)**: All global states are informationally complete.

3. **Semi-classical Validity (SCV)**: Semi-classical gravity is an effective theory that couples aspects of general relativity and quantum field theory; it's valid at most up to perturbative corrections in sub-Planckian regimes.

4. **Empirical Adequacy Condition (EAC)**: Empirically-confirmed predictions are recovered in their respective regimes.

5. **Universality Argument (UA)**: Black holes consist of Planckian degrees of freedom whose aggregate, statistical behavior recovers Bekenstein-Hawking entropy, which in turn reproduces universal thermodynamic phenomenology.

6. **Generalized Correspondence Principle (GCP)**: Effective theories are recovered in the appropriate limits.

Any attempt to evade information loss must involve modifying Hawking's original framework and/or introducing novel physics such that semi-classical gravity is recovered as a low-energy approximation. Physicists already have their work cut out for them, but Sisyphus's boulder becomes much heavier without consensus-building around the identity of black holes.

Questions Regarding the Identity of Black Holes:

1. What's the ontology underlying Bekenstein-Hawking degrees of freedom?
2. Is Bekenstein-Hawking entropy an information measure over the degrees of freedom of the horizon, interior, or exterior?
3. What's the relationship between the degrees of freedom of the horizon, interior and exterior?

Sisyphus's freedom can be bought if black hole evaporation is a thermodynamic process because guiding principles in black hole thermodynamics and statistical mechanics, namely the Universality Argument (UA) and a coherent reductionist story involving quantum gravitational degrees of freedom, are indispensable to the posing and resolving of the paradox. The taxonomy splitting up safehouse and evacuation solutions divulges the need to dig deeper into a comparative analysis of causal and holographic Bekenstein-Hawking entropy. Consequently, the black hole information loss paradox doubles as a theoretical laboratory, and if a particular interpretation of Bekenstein-Hawking entropy wins out with respect to UA while producing a resolution at all levels of nested paradoxes, then we have robust theoretical evidence in favor of it. The triumphant category will then be inserted into the "Holistic Desiderata to Transcend Black Hole Paradoxes".

Conclusion

For my doctoral dissertation, I have fruitfully gotten to the bottom of the black hole information loss paradox, which has been misleadingly characterized as indeterministic non-unitary evolution and clumsily parameterized as a paradigmatic clash between unitarity and the equivalence principle. 'Information' is the philosophically-loaded concept that has concomitantly revealed and obscured the radical tensions inherent in black hole evaporation.

My first order of clarification in Chapter 1 has been to decipher four connotations of information loss embedded in non-unitarity as 1) the Second Law of Thermodynamics, 2) indeterminism, 3) a variation in degrees of freedom, and 4) the spontaneous appearance of external entanglement, each of which corresponds to an appropriate entropic measure, while arguing that the latter two are relevant to the framing and resolution of the paradox.

My second order of clarification in Chapter 2 has been to expound how the relevant forms of information loss, associated with a decrease in maximal Boltzmann entropy and an increase in global von Neumann entropy respectively, engender what I've branded the "paradox of phantom entanglement". Black hole evaporation within Hawking's semi-classical framework insinuates that late-time Hawking radiation is an entangled global system, a contradiction in terms. Prospective solutions are then tasked with demonstrating how late-time Hawking radiation is either exclusively an entangled subsystem, in which a black hole remnant lingers as an information safehouse, or exclusively an unentangled global system, in which information is evacuated to the exterior.

My third order of clarification in Chapter 3 has been to expose the interpretation of Bekenstein-Hawking entropy – as an information measure over horizon states (safehouse solutions) versus nonlocalizable states (evacuation solutions) – in driving an additional point of contention. Insofar as Bekenstein-Hawking entropy underwrites black hole thermodynamics and statistical mechanics, its magnitude implies that black hole degrees of freedom are Planckian, which is why a theory of high-energy quantum gravity is indispensable to learning about the underlying ontology. A more thorough grasp on what black holes are made of would help us curb the space of proposals, but numerous live approaches are competing to mold our understanding of black holes as composite quantum gravitational systems.

Quite the contrary, black hole thermodynamics and statistical mechanics should take up the mantle as guides to a final theory of quantum gravity. Without knowing

anything more about Bekenstein-Hawking entropy, the most sensible place to look for clues is terrestrial thermodynamics and statistical mechanics, with unification as the goal. I've defined an original guiding principle, the Universality Argument (UA), which states that evaporating black holes fall into the same universality class as non-gravitational quantum statistical systems, for which thermodynamic phenomenology is recovered. Laying out specific criteria to discern relevant similarity to terrestrial applications, however, is a nontrivial task, so cashing out the alleged continuum is the crux of the controversy.

That's why it's important to test hypotheses even in a theoretical laboratory. Does Bekenstein-Hawking entropy quantify entanglement across the horizon? Does it play the role of Gibbs entropy, thereby providing an uncertainty measure over loopy horizon geometries? Or does it reflect a black hole's total Boltzmann entropy and enumerate stringy and braney constituents? If a particular interpretation – causal or holographic – of Bekenstein-Hawking entropy optimizes theoretical virtues, reduces black hole thermodynamics to statistical mechanics, and produces a resolution to paradox, then we've amassed theoretical evidence in favor of it and the supporting quantum gravity machinery. In other words, the interpretation of Bekenstein-Hawking entropy is the litmus test to vet the overpopulated proposal space.

The conclusion of this dissertation opens the door for trailblazing work to scrutinize causal and holographic interpretations of Bekenstein-Hawking entropy against the benchmark of UA. A systematic comparative analysis would divulge whether safe-house or evacuation solutions are more likely to undermine UA and attenuate the reductive link between black hole thermodynamics and black hole statistical mechanics, without which transcending the black hole information loss paradox would be a pyrrhic victory. The long-term goal is to discern the most palatable interpretation of Bekenstein-Hawking entropy and evaluate how it builds off of but also deviates from our current understanding of entropy and information in terrestrial physics. In doing so, I hope to develop a working account of the metaphysics of information and begin to dissolve the aforementioned culture clash, thereby providing timely and much needed momentum to developing a functional theory of quantum gravity.

Appendices

A The Extended General Relativistic Argument

The extended general relativistic argument can be summarized as follows:

1. Degrees of freedom constitute the global pre-evaporation system if and only if they're on a Cauchy surface prior to the final evaporation event, E.
2. Σ_2 is a Cauchy surface prior to E.
3. Therefore, the degrees of freedom of Σ_2 constitute the global pre-evaporation system.
4. Degrees of freedom are conserved from pre-to-post-evaporation if and only if they're on Σ_2 and Σ_3.
5. Black hole interior and Hawking radiation degrees of freedom are on Σ_2.
6. Hawking radiation degrees of freedom are on Σ_3.
7. Therefore, Hawking radiation degrees of freedom are conserved from pre-to-post-evaporation.
8. Degrees of freedom are annihilated from pre-to-post evaporation if and only if they're on Σ_2 but not Σ_3.
9. Black hole interior degrees of freedom are on Σ_2 but not Σ_3.
10. Therefore, black hole interior degrees of freedom are annihilated from pre-to-post-evaporation.
11. Degrees of freedom do not constitute the global post-evaporation system if and only if they are annihilated from pre-to-post-evaporation.
12. Therefore, black hole interior degrees of freedom do not constitute the global post-evaporation system.
13. Degrees of freedom constitute the global post-evaporation system if and only if they are conserved from pre-to-post-evaporation.

14. Therefore, Hawking radiation degrees of freedom constitute the global post-evaporation system.

Conclusion:

15. Therefore, the degrees of freedom of Σ_3 constitute the global post-evaporation system.

B The Extended Quantum Theoretic Argument

The extended quantum theoretic argument can be summarized as follows:

1. Σ_2 is a pre-evaporation vacuum state if and only if it contains entangled positive and negative-energy degrees of freedom forming Hawking pairs.
2. Σ_2 contains entangled positive and negative-energy degrees of freedom forming Hawking pairs.
3. Therefore, Σ_2 is a pre-evaporation vacuum state.
4. Positive-energy degrees of freedom of Σ_2 constitute an entangled pre-evaporation subsystem if and only if negative-energy degrees of freedom of Σ_2 are traced out.
5. Negative-energy degrees of freedom of Σ_2 are traced out.
6. Therefore, positive-energy degrees of freedom of Σ_2 constitute a pre-evaporation entangled subsystem.
7. If positive-energy degrees of freedom of Σ_2, constitute a pre-evaporation entangled subsystem, then they are also on a mixed state, Σ_{2+}.
8. Therefore, positive-energy degrees of freedom of Σ_2 are also on a mixed state, Σ_{2+}.
9. Degrees of freedom are conserved from pre-to-post-evaporation if and only if they're on Σ_{2+} and Σ_3.
10. State mixedness is preserved from pre-to-post-evaporation if and only if Σ_{2+} and Σ_3 are mixed states.
11. Positive-energy degrees of freedom of Σ_3 are also on a mixed state, Σ_3.
12. Therefore, positive-energy degrees of freedom are conserved from pre-to-post-evaporation.
13. Therefore, state mixedness is preserved from pre-to-post-evaporation.
14. The dynamical evolution from Σ_{2+} to Σ_3 is unitary if and only if degrees of freedom are conserved and state mixedness is preserved from pre-to-post-evaporation.

15. Therefore, the dynamical evolution from Σ_{2+} to Σ_3 is unitary.

16. Entanglement is preserved from pre-to-post-evaporation if and only if the dynamical evolution from Σ_{2+} to Σ_3 is unitary.

17. Therefore, entanglement is preserved from pre-to-post evaporation.

18. The degrees of freedom of Σ_3 constitute an entangled post-evaporation system if and only if entanglement is preserved from pre-to-post evaporation.

19. Therefore, the degrees of freedom of Σ_3 constitute an entangled post-evaporation system.

20. Systems are entangled if and only if they're subsystems.

Conclusion:

21. Therefore, the degrees of freedom of Σ_3 constitute an entangled post-evaporation subsystem.

Bibliography

Adlam, E. (2022, February). Laws of nature as constraints. *Foundations of Physics 52*(1).

Albert, D. Z. (1992). *Quantum Mechanics and Experience*. Cambridge Ma: Cambridge University Press.

Albert, D. Z. (2000). *Time and Chance*. Harvard University Press.

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

Almheiri, A., R. Mahajan, J. Maldacena, and Y. Zhao (2020, March). The Page curve of Hawking radiation from semiclassical geometry. *Journal of High Energy Physics 2020*(3).

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

Asaro, C. (1996). Complex speeds and special relativity. *American Journal of Physics 64*(4), 421–429.

Ashtekar, A., J. C. Baez, A. Corichi, and K. Krasnov (1998, February). Quantum geometry and black hole entropy. *Physical Review Letters 80*(5), 904–907.

Ashtekar, A. and P. Singh (2011, September). Loop quantum cosmology: a status report. *Classical and Quantum Gravity 28*(21), 213001.

Banks, T., L. Susskind, and M. E. Peskin (1984). Difficulties for the evolution of pure states into mixed states. *Nuclear Physics B 244*(1), 125–134.

Batterman, R. W. (2000). Multiple realizability and universality. *The British Journal for the Philosophy of Science 51*(1), 115–145.

Batterman, R. W. (2019). Universality and rg explanations. *Perspectives on Science 27*(1), 26–47.

Batterman, R. W. (2021). *A Middle Way: A Non-Fundamental Approach to Many-Body Physics*. Oxford University Press.

Bekenstein, J. D. (1973, Apr). Black holes and entropy. *Phys. Rev. D 7*(8), 2333–2346.

Bekenstein, J. D. (1974, June). Generalized second law of thermodynamics in black-hole physics. *Physical Review D 9*, 3292–3300.

Bekenstein, J. D. (1975). Statistical black-hole thermodynamics. *Phys. Rev. D 12*, 3077–3085.

Bekenstein, J. D. (1981). Energy cost of information transfer. *Physical Review Letters 46*(10), 623–626.

Bekenstein, J. D. (1994). Do we understand black hole entropy? https://arxiv.org/abs/gr-qc/9409015v2.

Bekenstein, J. D. (2001). The limits of information. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 32*(4), 511–524.

Bekenstein, J. D. (2007). Information in the holographic universe. *Scientific American*.

Bekenstein, J. D. and M. Schiffer (1990). Quantum limitations on the storage and transmission of information. *International Journal of Modern Physics C 1*(4), 355–422.

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

Bianchi, E., M. Christodoulou, F. D'Ambrosio, H. M. Haggard, and C. Rovelli (2018, October). White holes as remnants: A surprising scenario for the end of a black hole. *Classical and Quantum Gravity 35*(22), 225003.

Bokulich, P. (2003). *Horizons of Description: Black Holes and Complementarity*. Ph. D. thesis, University of Notre Dame.

Bokulich, P. (2011). Interactions and the consistency of black hole complementarity. *International studies in the philosophy of science 25*(4), 371–386.

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

Bousso, R. (2013, June). Complementarity is not enough. *Physical Review D 87*(12).

Brillouin, L. (2013). *Science and Information Theory* (2 ed.). Dover Publications, Inc.

Calmet, X., R. Casadio, S. D. Hsu, and F. Kuipers (2022, March). Quantum hair from gravity. *Physical Review Letters 128*(11).

Calosi, C. and M. Morganti (2021). Interpreting quantum entanglement: Steps towards coherentist quantum mechanics. *The British Journal for the Philosophy of Science 72*(3), 865–891.

Carlip, S. (2014, October). Black hole thermodynamics. *International Journal of Modern Physics D 23*(11), 1430023.

Casadio, R. and B. Harms (2011). Microcanonical description of (micro) black holes. *Entropy 13*(2), 502–517.

Chandrasekaran, V., G. Penington, and E. Witten (2023, April). Large N algebras and generalized entropy. *Journal of High Energy Physics 2023*(4).

Chatterjee, S., M. Parikh, and S. Sarkar (2012, January). The black hole membrane paradigm in f(R) gravity. *Classical and Quantum Gravity 29*(3), 035014.

Chen, E. K. (2020). Time's arrow in a quantum universe: On the status of statistical mechanical probabilities. In V. Allori (Ed.), *Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature*, pp. 479–515. World Scientific.

Chen, P., Y. C. Ong, and D.-H. Yeom (2015). Black hole remnants and the information loss paradox. *Physics Reports 603*, 1–45.

Christodoulou, M. and C. Rovelli (2015, March). How big is a black hole? *Physical Review D 91*(6).

Clifton, R. and H. Halvorson (2001). Entanglement and open systems in algebraic quantum field theory. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 32*(1), 1–31.

Crowther, K. (2018a). Defining a crisis: the roles of principles in the search for a theory of quantum gravity. *Synthese*.

Crowther, K. (2018b). Inter-theory relations in quantum gravity: Correspondence, reduction and emergence. *Studies in History and Philosophy of Modern Physics 63*, 74–85.

Crowther, K. and S. de Haro (2022). Four attitudes towards singularities in the search for a theory of quantum gravity. In A. Vassallo (Ed.), *The Foundations of Spacetime Physics: Philosophical Perspectives*, Chapter 9, pp. 223 – 250. Taylor Francis Group.

Crowther, K., N. Linnemann, and C. Wüthrich (2021, 07). What we cannot learn from analogue experiments. *Synthese 198*, 3701–3726.

Crowther, K., N. S. Linnemann, and C. Wüthrich (2021). What we cannot learn from analogue experiments. *Synthese 198*, 3701–3726.

Curiel, E. (2014). Classical black holes are hot. https://arxiv.org/pdf/1408.3691.pdf.

Curiel, E. (2020). Models of black hole evaporation, or, what to do when you can't solve equations. Video lecture. https://youtu.be/hvy9lYLptBw?si=kou95mAZ0V7JED-q.

Curiel, E. (2021). Kinematics, dynamics, and the structure of physical theory. http://strangebeautiful.com/papers/curiel-kins-dyns-struc-theory.pdf.

Curiel, E. (2023a). Energy, entropy and the intimate relations between the two in semi-classical gravity. Video lecture. https://youtu.be/9j21XgmIckw?si=NgEGr44uMoHPyJfU.

Curiel, E. (2023b). The Hawking effect, its desiderata and its discontents. Video lecture. https://youtu.be/jFZ2HSkMTvY?si=VOiaKau9ykWYd53O.

Dawid, R. (2019). The significance of non-empirical confirmation in fundamental physics. In R. Dardashti, R. Dawid, and K. Thebault (Eds.), *Why Trust a Theory? Epistemology of ModernPhysics*, pp. 99–119. Cambridge University Press.

Dougherty, J. and C. Callender (2016). Black hole thermodynamics: More than an analogy? http://philsci-archive.pitt.edu/13195/.

Earman, J. (1996). Tolerance for spacetime singularities. *Foundations of Physics 26*(5), 623–640.

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

El Skaf, R. and P. Palacios (2022). What can we learn (and not learn) from thought experiments in black hole thermodynamics? *Synthese 200*(6).

Engelhardt, N. and Å. Folkestad (2022, jul). Negative complexity of formation: the compact dimensions strike back. *Journal of High Energy Physics 2022*(7).

Engelhardt, N. and A. C. Wall (2015, January). Quantum extremal surfaces: Holographic entanglement entropy beyond the classical regime. *Journal of High Energy Physics 2015*(1).

Engelhardt, N. and A. C. Wall (2017, April). No simple dual to the causal holographic information? *Journal of High Energy Physics 2017*(4).

Esfeld, M. (2004). Quantum entanglement and a metaphysics of relations. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 35*(4), 601–617.

Floridi, L. (2004). Open problems in the philosophy of information. *Metaphilosophy 35*(4), 554–582.

Fraser, D. (2022). Particles in quantum field theory. In E. Knox and A. Wilson (Eds.), *The Routledge Companion to Philosophy of Physics*, pp. 323–336. Routledge.

Fredenhagen, K. and R. Haag (1990). On the derivation of hawking radiation associated with the formation of a black hole. *Communications in mathematical physics 127*(2), 273–284.

French, S. and K. McKenzie (2012). Thinking outside the toolbox: Towards a more productive engagement between metaphysics and philosophy of physics. *European Journal of Analytic Philosophy 8*(1), 42–59.

Gambini, R. and J. Pullin (2008, March). Holography from loop quantum gravity. *International Journal of Modern Physics D 17*, 545–549.

Giddings, S. B. (1992, August). Black holes and massive remnants. *Physical Review D 46*(4), 1347–1352.

Großardt, A. (2022, February). Three little paradoxes: Making sense of semiclassical gravity. *AVS Quantum Science 4*(1).

Ha, Y. K. (2003, November). The gravitational energy of a black hole. *General Relativity and Gravitation 35*, 2045–2050.

Harlow, D. and P. Hayden (2013). Quantum computation vs. firewalls. *Journal of High Energy Physics 2013*(6).

Hawking, S. W. (1975). Particle creation in black holes. *Communications in Mathematical Physics 43*(6), 199–220.

Hawking, S. W. (1976, November). Breakdown of predictability in gravitational collapse. *Phys. Rev. D 14*, 2460–2473.

Hawking, S. W., M. J. Perry, and A. Strominger (2016, June). Soft hair on black holes. *Physical Review Letters 116*(23).

Hayden, P. and J. Preskill (2007, September). Black holes as mirrors: Quantum information in random subsystems. *Journal of High Energy Physics 2007*(09), 120–138.

Hossenfelder, S. (2004). What black holes can teach us.

Hossenfelder, S. (2012). Comment on the black hole firewall.

Hossenfelder, S. (2018). *Lost in Math: How Beauty Leads Physics Astray*. New York: Basic Books.

Hossenfelder, S. and L. Smolin (2010, March). Conservative solutions to the black hole information problem. *Physical Review D 81*(6).

Hsu, S. D. (2007, January). Spacetime topology change and black hole information. *Physics Letters B 644*(1), 67–71.

Hubeny, V. E. and M. Rangamani (2012, June). Causal holographic information. *Journal of High Energy Physics 2012*(6).

Huggett, N. and K. Matsubara (2021, July). Lost horizon? – Modeling black holes in string theory. *European Journal for Philosophy of Science 11*(3).

Jacobson, T. (1995, August). Thermodynamics of spacetime: The Einstein equation of state. *Phys. Rev. Lett. 75*, 1260–1263.

Jacobson, T. (1999). On the nature of black hole entropy. *Eighth Canadian conference on general relativity and relativistic astrophysics*.

Jacobson, T., D. Marolf, and C. Rovelli (2005). Black hole entropy: Inside or out? *International Journal of Theoretical Physics 44*, 1807–1837.

Jaksland, R. (2021). Entanglement as the world-making relation: Distance from entanglement. *Synthese 198*, 9661–9693.

Jaksland, R. (2023). A trilemma for naturalized metaphysics. *Ratio 36*(1), 1–10.

Jaynes, E. T. (1957a). Information theory and statistical mechanics. *Physical Review 106*(4), 620–630.

Jaynes, E. T. (1957b). Information theory and statistical mechanics. ii. *Physical Review 108*(2), 171–190.

Jusufi, K., E. Moulay, J. Mureika, and A. F. Ali (2023, April). Einstein-Rosen bridge from the minimal length. *The European Physical Journal C 83*(4).

Kiem, Y., H. Verlinde, and E. Verlinde (1995, dec). Black hole horizons and complementarity. *Physical Review D 52*(12), 7053–7065.

Ladyman, J. and D. Ross (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford University Press.

Landauer, R. (1996). The physical nature of information. *Physics Letters A 217*(4), 188–193.

Lesourd, M. (2018, December). Causal structure of evaporating black holes. *Classical and Quantum Gravity 36*(2), 025007.

Linnemann, N. S. and M. R. Visser (2018). Hints towards the emergent nature of gravity. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 64*, 1–13.

Lowe, D. A., J. Polchinski, L. Susskind, L. Thorlacius, and J. Uglum (1995, Dec). Black hole complementarity versus locality. *Physical Review D 52*(12), 6997–7010.

Malament, D. B. (2012). *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory*. Chicago: Chicago University 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, August). Cool horizons for entangled black holes. *Fortschritte der Physik 61*(9), 781–811.

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. (2011). What is a physically reasonable space-time? *Philosophy of science 78*(3), 410–420.

Mann, R. M. (2015). *Black Holes: Thermodynamics, Information, and Firewalls*. Springer International Publishing.

Marolf, D. (2009, February). Black holes, AdS, and CFTs. *General Relativity and Gravitation 41*(4), 903–917.

Maroney, O. J. E. and C. G. Timpson (2017, April). How is there a physics of information? On characterising physical evolution as information processing.

Mathur, S. D. (2009). The information paradox: a pedagogical introduction. *Classical and Quantum Gravity 26*(22), 224001.

Mathur, S. D. and D. Turton (2014, January). Comments on black holes I: the possibility of complementarity. *Journal of High Energy Physics 2014*(1).

Matsuo, Y. (2021, July). Islands and stretched horizon. *Journal of High Energy Physics 2021*(7).

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

Mattingly, J. (2009). Mongrel gravity. *Erkenntnis (1975-) 70*(3), 379–395.

Maudlin, T. (2002). *Quantum Non-locality and Relativity* (2 ed.). Oxford: B. Blackwell.

Maudlin, T. (2017). (Information) paradox lost.

Maudlin, T. (2018). Ontological clarity via canonical presentation: Electromagnetism and the Aharonov–Bohm effect. *Entropy 20*(6), 465–.

Maudlin, T., E. Okon, and D. Sudarsky (2020). On the status of conservation laws in physics: Implications for semiclassical gravity. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 69*, 67–81.

Mermin, N. D. (1998, September). What is quantum mechanics trying to tell us? *American Journal of Physics 66*(9), 753–767.

Myrvold, W. C. (2021). *Beyond Chance and Credence: A Theory of Hybrid Probabilities*. Oxford University Press.

Okon, E. and D. Sudarsky (2014). Benefits of objective collapse models for cosmology and quantum gravity. *Foundations of Physics 44*(2), 114–143.

Okon, E. and D. Sudarsky (2017, January). Black holes, information loss and the measurement problem. *Foundations of Physics 47*(1), 120–131.

Osuga, K. and D. N. Page (2018, March). Qubit transport model for unitary black hole evaporation without firewalls. *Physical Review D 97*(6).

Page, D. N. (1980, Feb). Is black-hole evaporation predictable? *Phys. Rev. Lett. 44*, 301–304.

Page, D. N. (1993a). Average entropy of a subsystem. *Physical Review Letters 71*(9), 1291–1294.

Page, D. N. (1993b, December). Information in black hole radiation. *Physical Review Letters 71*(23), 3743–3746.

Page, D. N. (1995). Black hole information.

Page, D. N. (2013, September). Time dependence of Hawking radiation entropy. *Journal of Cosmology and Astroparticle Physics 2013*(09), 028–028.

Parikh, M. K. and F. Wilzcek (2000, January). Hawking radiation as tunneling. *Physical Review Letters 85*, 5042–5045.

Penington, G. (2020, September). Entanglement wedge reconstruction and the information paradox. *Journal of High Energy Physics* (9).

Perez, A. (2017, October). Black holes in loop quantum gravity. *Reports on Progress in Physics 80*(12), 126901.

Polchinski, J. (2017). The black hole information problem. In *New Frontiers in Fields and Strings*, Chapter 6, pp. 353–397.

Polchinski, J. and A. Strominger (1994, December). Possible resolution of the black hole information puzzle. *Physical Review D 50*(12), 7403–7409.

Preskill, J. (1992). Do black holes destroy information?

Prunkl, C. E. A. and C. G. Timpson (2019). Black hole entropy is thermodynamic entropy.

Raju, S. (2022). Lessons from the information paradox. *Physics Reports 943*, 1–80. Lessons from the information paradox.

Rovelli, C. (1996, October). Black hole entropy from loop quantum gravity. *Physical Review Letters 77*(16), 3288–3291.

Rovelli, C. (2019, August). The subtle unphysical hypothesis of the firewall theorem. *Entropy 21*(9), 839.

Ruetsche, L. (2011). *Interpreting Quantum Theories: The Art of the Possible*. Oxford University Press.

Ryu, S. and T. Takayanagi (2006). Holographic derivation of entanglement entropy from the anti–de Sitter space/conformal field theory correspondence. *Physical Review Letters 96*(18).

Sainsbury, R. M. (2009). *Paradoxes* (3 ed.). Cambridge University Press.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal 27*(3), 379–423, 623–656.

Sheldon Goldstein, Joel L. Lebowitz, R. T. N. Z. (2020, May). In V. Allori (Ed.), *Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature*, Chapter 14, pp. 519–581. World Scientific.

Shenker, O. (2019). Information vs. entropy vs. probability. *European Journal for Philosophy of Science 10*(1), 1–25.

Sonner, J. and M. Vielma (2017, November). Eigenstate thermalization in the Sachdev-Ye-Kitaev model. *Journal of High Energy Physics 2017*(11).

Sorkin, R. D. (1997). The statistical mechanics of black hole thermodynamics.

Sorkin, R. D. (2011). Ten theses on black hole entropy.

Strominger, A. and C. Vafa (1996, June). Microscopic origin of the Bekenstein-Hawking entropy. *Physics Letters B 379*(1-4), 99–104.

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

Susskind, L. (2008). *The Black Hole War*. Little, Brown and Company.

Susskind, L. (2012). The transfer of entanglement: The case for firewalls.

Susskind, L. (2013). Black hole complementarity and the Harlow-Hayden conjecture.

Susskind, L. and J. Lindesay (2004). *An Introduction to Black Holes, Information and the String Theory Revolution: The Holographic Universe*. World Scientific Publishing Co Pte Ltd.

Susskind, L. and L. Thorlacius (1994, Jan). Gedanken experiments involving black holes. *Physical Review D 49*(2), 966–974.

Susskind, L., L. Thorlacius, and J. Uglum (1993, Oct). The stretched horizon and black hole complementarity. *Phys. Rev. D 48*, 3743–3761.

Tahko, T. E. (2018). Fundamentality. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 ed.). Metaphysics Research Lab, Stanford University.

Teufel, S. (2022). Boltzmann entropy in quantum mechanics. In *The Chimera of Entropy II: Arrow of Time*.

Thébault, K. P. Y. (2019). What can we learn from analogue experiments? In *Why Trust a Theory? Epistemology of Fundamental Physics*, pp. 184–201. Cambridge University Press.

t'Hooft, G. (1996, October). The scattering matrix approach for the quantum black hole: An overview. *International Journal of Modern Physics A 11*(26), 4623–4688.

Unruh, W. G. and R. M. Wald (2017). Information loss. *Reports on Progress in Physics 80*(9), 092002.

van Dongen, J. and S. de Haro (2004). On black hole complementarity. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 35*(3), 509–525.

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

Wald, R. M. (2001, July). The thermodynamics of black holes. *Living Reviews in Relativity 4*(1).

Wald, R. M. (2019, September). Particle and energy cost of entanglement of Hawking radiation with the final vacuum state. *Physical Review D 100*(6).

Wallace, D. (2018). The case for black hole thermodynamics part I: Phenomenological thermodynamics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 64*, 52–67.

Wallace, D. (2019). The case for black hole thermodynamics part II: Statistical mechanics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 66*, 103–117.

Wallace, D. (2020). Why black hole information loss is paradoxical. In N. Huggett, K. Matsubara, and C. Wüthrich (Eds.), *Beyond Spacetime: The Foundations of Quantum Gravity*, pp. 209–236. Cambridge University Press.

Wallace, D. (2022). Quantum gravity at low energies. *Studies in History and Philosophy of Science Part A 94*(C), 31–46.

Wallace, D. (2023). The logic of the past hypothesis. In B. Loewer, B. Weslake, and E. B. Winsberg (Eds.), *The Probability Map of the Universe: Essays on David Albert's* Time and Chance. Cambridge MA: Harvard University Press.

Wheeler, J. A. (2002). Information, physics, quantum: The search for links. In A. Hey (Ed.), *Feynman and Computation*, Chapter 19, pp. 309–336. Boca Raton: CRC Press.

Witten, E. (1998). Anti-de Sitter space and holography. *Advances in Theoretical and Mathematical Physics 2*, 253–291.

Wüthrich, C. (2005). To quantize or not to quantize: Fact and folklore in quantum gravity. *Philosophy of Science 72*(5), 777–788.

Wüthrich, C. (2019). Are black holes about information? In R. Dawid, R. Dardashti, and K. Thébault (Eds.), *Why Trust a Theory? Epistemology of Fundamental Physics*, pp. 202–223. Cambridge, U.K.: Cambridge University Press.

Wüthrich, C. (2021). Quantum gravity from general relativity. In E. Knox and A. Wilson (Eds.), *The Routledge Companion to Philosophy of Physics*. Routledge.

Zurek, W. H. (1982, Dec). Entropy evaporated by a black hole. *Phys. Rev. Lett. 49*, 1683–1686.