Foundational puzzles on quantum universality:

From Wigner's friend to black holes

Laurens Walleghem

PhD

University of York
Mathematics

March 2026

Abstract

Despite its remarkable empirical success and predictive power, quantum theory is riddled with mysteries. One such conceptual puzzle is the measurement problem: why do quantum systems in superposition collapse to a single-valued outcome upon being measured? The scenario of Wigner's friend illustrates this tension, with recent extensions deriving strong no-go results for the nature of reality. Another fundamental puzzle arises at the interface of quantum theory and general relativity—the black hole information paradox. This thesis further explores these two foundational paradoxes.

Part I of this thesis focuses on extensions of the Wigner's friend scenario. We introduce several novel extended Wigner's friend arguments, and establish connections with notions of nonclassicality, i.e. the study of quantum phenomena that evade classical explanation. Key results are a refined version of the Frauchiger–Renner paradox, new extended Wigner's friend scenarios based on contextuality, and an argument against the absoluteness of free choices based on the Pusey–Barrett–Rudolph theorem. The latter result complements the Local Friendliness no-go theorem, which challenged the absoluteness of experimentally obtained measurement outcomes. In Part II of this thesis, we turn to black hole puzzles and the interplay of quantum theory and gravity. We present a paradox that unifies extended Wigner's friend arguments with black hole puzzles, strengthening the cloning and firewall paradoxes. After discussing subtleties in these paradoxes, we next present an analysis combining elements of quantum theory with gravity, analyzing the detector response near a superposed black hole using elements of quantum reference frames.

Contents

Abstract **2**

Contents **3**

List of Tables **9**

List of Figures **9**

Acknowledgements **12**

Author's declaration **14**

Abbreviations **16**

1 Introduction **17**

1.1 Part I: Universality in quantum foundations 18

1.1.1 The measurement problem, a universality hypothesis and Wigner's friend 18

1.1.2 Part I: Contributions 19

1.2 Part II: Universality and black holes 21

1.2.1 The black hole information problem and related paradoxes 21

1.2.2 Part II: Contributions 23

1.3 Structure of thesis 24

I Universality in quantum foundations 26

2 Quantum preliminaries 28

2.1 A short recap of quantum theory . . . 28
2.1.1 Notation in linear algebra . . . 28
2.1.2 Quantum theory postulates . . . 29
2.1.3 Remarks . . . 30
2.2 Notions of nonclassicality . . . 34
2.2.1 Kochen–Specker contextuality and Bell nonlocality . . . 36
2.2.2 Generalized noncontextuality . . . 41
2.3 The measurement problem, Wigner's friend and extensions . . . 44
2.3.1 Wigner's friend scenario . . . 45
2.4 The Frauchiger–Renner paradox . . . 48
2.4.1 Entanglement version of the FR paradox . . . 48
2.4.2 A slightly stronger modified FR paradox . . . 51
2.4.3 Logical non-locality powers the FR paradox . . . 51
2.5 The Local Friendliness no-go theorem . . . 54
2.5.1 Local Friendliness scenarios . . . 55
2.5.2 No-go theorem . . . 56

3 A refined Frauchiger–Renner paradox based on strong contextuality 59

3.1 Refining the Frauchiger–Renner paradox . . . 60
3.1.1 Strong contextuality: the GHZ–Mermin argument . . . 60
3.1.2 Description of the GHZ–FR paradox . . . 61
3.2 Two GHZ–FR no-go theorems . . . 65
3.2.1 GHZ–FR truth no-go theorem . . . 66
3.2.2 GHZ–FR agreement no-go theorem . . . 73
3.2.3 A possible resolution . . . 78
3.3 Comparison with the FR paradox . . . 80
3.3.1 Why the GHZ–FR paradox is stronger than the FR paradox . . . 80
3.4 Related work on the FR paradox . . . 83
3.4.1 Previous responses to the FR paradox . . . 84
3.4.2 Related work on the GHZ–FR paradox . . . 87
3.4.3 Relation to the Local Friendliness no-go theorem . . . 88
3.5 Conclusion . . . 89

3.5.1 Summary . . . 89
3.5.2 Outlook . . . 90

4 Extended Wigner's friend arguments and noncontextuality 92
4.1 An extended Wigner's friend paradox based on KS contextuality . . . 94
4.1.1 Assumptions . . . 94
4.1.2 The 5-cycle noncontextuality scenario . . . 95
4.1.3 An EWF scenario with measurements on a single system . . . 96
4.1.4 No-go theorem . . . 97
4.1.5 Discussion and outlook . . . 100
4.2 Deriving Local Friendliness inequalities from KSNC constraints . . . 102
4.2.1 Technical consequences of the Local Friendliness assumptions for LF scenarios . . . 102
4.2.2 The compatibility graph for a N_A, N_B-setting LF scenario . . . 105
4.2.3 An alternative expression of the LF constraints . . . 107
4.2.4 The 2,2-setting LF scenario . . . 111
4.2.5 The 3,2-setting LF scenario . . . 111
4.2.6 The 3,3-setting LF scenario . . . 114
4.2.7 Generalizations to multipartite sequential LF scenarios . . . 116
4.3 Constructing new LF scenarios from KSNC arguments . . . 117
4.3.1 The 5-cycle example . . . 117
4.3.2 The Peres–Mermin example . . . 123
4.3.3 Translating any possibilistic KSNC argument into a LF argument 129
4.4 An extended Wigner's friend paradox based on generalized contextuality . 129
4.4.1 Bell nonlocality and generalized contextuality . . . 130
4.4.2 Rephrasing the Local Agency assumption . . . 132
4.4.3 Noncontextual Friendliness no-go theorem . . . 134
4.4.4 Discussion . . . 138

5 Are free choices absolute, when Pusey–Barrett–Rudolph are friends of Wigner? 141
5.1 The Pusey–Barrett–Rudolph argument . . . 141
5.2 Are free choices absolute? . . . 143
5.3 Generalizing to arbitrary choice amplitudes . . . 147
5.4 On the modeling of Alice's and Bob's choices . . . 148

5.5 Use of correlations . . . 148
5.6 Discussion . . . 149

6 Intermezzo: the Sleeping Beauty problem 151
6.1 Introduction . . . 152
6.2 A Thirder's Tale of Beauty and the Prince . . . 153
6.3 A variation of the Sleeping Beauty problem . . . 156
6.4 Relating to literature on Sleeping Beauty . . . 158
6.4.1 Elga's argument for 1/3 and the Principle of Indifference . . . 159
6.4.2 The Thirders vs. the Halfers . . . 160
6.5 Conclusion . . . 161

II Universality and black holes 162

7 A brief introduction to black hole puzzles 164
7.1 Literature on the black hole information puzzle . . . 165
7.2 A black hole's soft hair . . . 166
7.3 What about paradoxes for infalling observers? . . . 168

8 Wigner's friend's black hole adventure 169
8.1 Combining Wigner's friend and black hole cloning . . . 170
8.1.1 Critical note about black hole assumptions . . . 171
8.1.2 Protocol and no-go theorem . . . 172
8.1.3 Do the outcome records of observers inside get altered by operations outside? . . . 175
8.1.4 Comparison with the no-cloning argument . . . 176
8.1.5 Combining other extended Wigner's friend paradoxes with the cloning paradox . . . 177
8.2 Combining Wigner's friend and the firewall argument . . . 177
8.2.1 Preliminaries: the firewall argument . . . 178
8.2.2 Protocol and no-go theorem . . . 180
8.3 Relating to the work by Hausmann and Renner . . . 182
8.4 Discussion . . . 184
8.4.1 Subtleties in black hole paradoxes . . . 185
8.4.2 Implications and future directions . . . 189

9 Particle detector near position-superposed black hole 191
9.1 Motivation: bottom-up quantum gravity, quantum reference frames and detectors . . . 192
9.2 Introducing BTZ black holes, Unruh–DeWitt detectors and quantum reference frames . . . 193
9.2.1 BTZ Black hole . . . 193
9.2.2 Unruh–DeWitt detectors . . . 196
9.2.3 Quantum Reference Frames . . . 199
9.3 Set-up, QRF transformation and reference fields . . . 202
9.3.1 Setup . . . 202
9.3.2 A treatment of the position-superposition using reference fields . 204
9.4 Detector response near position-superposed black hole . . . 212
9.4.1 Results . . . 212
9.4.2 Numerical analysis . . . 214
9.4.3 Mass- vs. position-superposition . . . 215
9.5 Discussion and outlook . . . 218

10 Conclusions and future work 220
10.1 Conclusions . . . 220
10.2 Part I: discussion and outlook . . . 220
10.3 Part II: discussion and outlook . . . 225

A Extra black hole protocols of Chapter 8 228
A.1 A Local Friendliness version of the first protocol . . . 228
A.2 No-signalling firewall argument . . . 232

B Appendix for Chapter 9 234
B.1 The BTZ black hole as a quotient of AdS_3 . . . 234
B.2 BTZ correlation functions . . . 235
B.3 Interaction picture calculation . . . 237
B.4 Calculating integrals for the measurement probabilities . . . 241
B.4.1 Calculating the cross-term $\mathcal{F}_{12}(\Omega)$ . . . 241
B.4.2 Calculating $\mathcal{F}_J(E)$ . . . 243
B.4.3 Relating to the coordinates of Ref. [119] and final expressions . . 245
B.5 Derivation of interaction Hamiltonian for superposed black hole . . . 247
B.6 Derivation of the analytical expressions of Section 9.4.3 . . . 248

B.6.1 Summary of expressions 248
B.6.2 Analytic expressions for single term 250
B.6.3 Analytic expressions for interference term 252

References **257**

List of Tables

2.1 (Im)possibility table for the Hardy model 54

3.1 (Im)possibility table for the GHZ–Mermin scenario 62
3.2 Assumptions of the GHZ–FR truth no-go theorem 89
3.3 Assumptions of the GHZ–FR agreement no-go theorem 90

List of Figures

2.1 n-cycle compatibility hypergraphs 37
2.2 Simplest nontrivial scenario for generalized contextuality 44
2.3 Wigner's friend scenario 46
2.4 Sketch of the FR paradox 50
2.5 Mapping a Bell nonlocality scenario to an FR scenario 52
2.6 Local Friendliness setup 56

3.1 The GHZ–FR protocol 63
3.2 Possible lightcone structures for the GHZ–FR protocol 64

3.3 An example of a reasoning situation following the GHZ–FR protocol that is not allowed by possibilistic Born rule 67
3.4 Assumptions of the GHZ–FR no-go theorems 71
3.5 Comparing the assumptions of the FR and GHZ–FR paradoxes 84

4.1 The compatibility graph of the 5-cycle scenario 96
4.2 Schematic depiction of an EWF scenario on a single system using 5-cycle contextuality 98
4.3 Compatibility graph for a N_A, N_B-setting LF scenario 105
4.4 The compatibility graph for A_1, A_i, B_1 and B_j 109
4.5 The compatibility graph for the 2,2-setting LF scenario 111
4.6 The compatibility graph of the 3,2-setting LF scenario 112
4.7 Special cases of the compatibility graph in Fig. 4.3 114
4.8 Subgraphs of the compatibility graph of the 3,3-setting LF scenario 115
4.9 The schematic representation of the LF 5-cycle scenario 118
4.10 Cartoon representation of the LF 5-cycle scenario 118
4.11 A compatibility graph for the measurements in the LF 5-cycle protocol . . . 120
4.12 The schematic proof using the LF 5-cycle protocol for the LF no-go theorem 121
4.13 Schematic diagram of the LF Peres–Mermin scenario 125
4.14 Cartoon representation of the LF Peres–Mermin scenario 127
4.15 From Bell nonlocality to prepare-and-measure scenarios 131
4.16 Simplest nontrivial scenario for generalized contextuality 132
4.17 Local Friendliness set-up 133
4.18 Noncontextual Friendliness setup 135

5.1 Combining Wigner's friend with the Pusey–Barrett–Rudolph argument . . . 144

6.1 Prince Probability and Sleeping Beauty 155

8.1 Combining extended Wigner's friend with black hole cloning 170
8.2 Combining extended Wigner's friend with the firewall argument 182
8.3 Continuing the spacetime post-evaporation 187
8.4 A critique of the cloning and firewall paradoxes 188

9.1 Maximally extended Penrose diagram of the BTZ black hole 194
9.2 Action of a quantum reference frame transformation 201
9.3 Set-up: detector near position-superposed black hole 202

9.4 Interferometric and detector response measurement probabilities for position-superposed black hole 214
9.5 Interferometric and detector response measurement probabilities for mass-superposed black hole [119] 216

A.1 Combining Local Friendliness with black hole cloning 229

Acknowledgements

This work would not have been possible without the help of many. I would like to thank my supervisors Dr. Matt Pusey and Dr. Stefan Weigert for their guidance, especially for our interesting conversations on the Frauchiger–Renner paradox and our refinement thereof, and for teaching me how to present scientific results, and Prof. Chris Fewster and Roger Colbeck for their TAP support. A special thanks goes to Dr. Rui Soares Barbosa; much of the work in the first part of my PhD arose from initial conversations with Rui. I would also like to thank my examiners Prof. Chris Fewster and Prof. Časlav Brukner for their useful comments, which significantly improved this thesis. I thank my collaborators Rafael Wagner, Yîlè Yīng, David Schmid, Lorenzo Catani, Rutvij Bhavsar, Shashaank Khanna and Carlo Cepollaro, all of whom I learned a great deal from. I would like to thank Raman Choudhary for the many interesting conversations on notions of nonclassicality, of crucial importance for my understanding thereof. I thank Dr. Benito Juárez–Aubry and Prof. Bernard Kay for many interesting discussions on black hole physics, and Prof. Chris Fewster for his always-helpful replies to my questions on (A)QFT. I thank Harkan Kirk-Karakaya for teaching me how to run programs on the UoY server. I would like to thank Prof. Nicolò Defenu for furthering my understanding of quantum field theory and his astonishing intuitive understanding of QFT (condensed matter) systems. I have truly learned a lot from Nicolò. I would like to thank Claudio Dappiaggi for helpful conversations on black hole physics and its viewpoint as a scattering problem. I thank Dr. Ricardo Faleiro and Prof. Časlav Brukner for interesting discussions on the firewall paradox, and Prof. David Mattingly for interesting discussions and exchanges on gravitational shielding and gauge invariance regarding quantum paradoxes. I thank Dr. Ambroise Grau for his template for this thesis.

There is one person I have not mentioned above in my list of collaborators, and that

is David Maibach, my partner in black hole crimes. Throughout my PhD, we have been having continuous conversations on the physics of black holes, which have advanced my understanding significantly.

Finally, I would like to express my sincere gratitude to my circle of family and friends, and all the new friends I made on the way. My parents, and sister, for their continued support, and without whom my whole education would not have been possible. Of the long list of my friends, I would like to thank Toon, Raman, Manisha, Quinten, Arne, Arno, Tanguy, Siebe, David, Jasper, Thomas, Yannick, Laurens, Michael, Lidewij, Marie, Joppe, João, Carolina, Anita, Davide, Angelos, Jaymin, Luisa, Osvaldo, Diogo, Eduardo, Toni, Leonardo, Lorenzo, Marco, Rafael, Timo, Nelson, Dieter, Kobe, Arne, Vincent, Jef, Bram, Will, Ari, Santiago, Alessandro, Baptiste, Tara, Justin, Zara, Gabriel, Zane, Máté, Rutvij, Peter, Shashaank, Kuntal, Filippo, Harkan, Vincenzo, Lewis, Ricky, Vasilis, Alistair, all the Bens and Sams in the office, Philip, Simen, Ross, Jamie, Sol, Dorsa, Ambroise, Ding, Alice, Simon, Samuel, Cody, Levent, Marie-Lou, Thomas, Victor, Tina, Lodewijk, Robin, Wilfried, Walter, Olivier, Marthe, Jan, Benjamin, Seba, David, Olivier L., Andreas V., and Khadija, for always supporting me, as well as those who I have not listed here. They have supported me almost as much as Zane has continuously supported United, even in the most desperate of times.

I acknowledge support from the United Kingdom Engineering and Physical Sciences Research Council (EPSRC) through the DTP Studentship EP/W524657/1, and thank INL and its QLOC group in Braga, Portugal for the kind hospitality.

Laurens Walleghem,
August 2026.

Author's declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

The results in Part I of this thesis are based on the following works, in the order of appearance:

Chapter 3:

L. Walleghem, R. S. Barbosa, M. Pusey and S. Weigert. A refined Frauchiger–Renner paradox based on strong contextuality, Quantum **10**, 2116 (2026).

Chapter 4:

L. Walleghem, R. Wagner, Y. Yīng and D. Schmid. Extended Wigner's friend paradoxes do not require nonlocal correlations, Phys. Rev. A **112**, 022212 (2025).

L. Walleghem, R. Wagner, Y. Yīng and D. Schmid. Connecting extended Wigner's friend arguments and noncontextuality, Quantum **9**, 1819 (2025).

L. Walleghem and L. Catani. An extended Wigner's friend no-go theorem inspired by generalized contextuality, arXiv:2502.02461 (2025).

Chapter 5:

L. Walleghem. Are free choices absolute, when internalized in Wigner's friend?, arXiv:2605.14538 (2026).

Chapter 6:

L. Walleghem. Stunned by Sleeping Beauty: How Prince Probability updates his forecast upon their fateful encounter, arXiv:2408.06797 (2024).

The results of Part II are based on the following works:

Chapter 8:

L. Walleghem. Wigner's friend's black hole adventure: an argument for complementarity?, arXiv:2507.05369 (2025).

Chapter 9:

L. Walleghem and C. Cepollaro. Particle detector near a position-superposed black hole. arXiv:2604.11897 (2026).

The following publication was left out of this thesis:

L. Walleghem, S. Khanna and R. Bhavsar. Comment on a No-Go Theorem for ψ-Ontic Models, Found Phys **55**, 23 (2025).

For proofreading parts of this work the online tools Grammarly and ChatGPT were used. I also acknowledge the use of ChatGPT (developed by OpenAI) and Claude (developed by Anthropic) as search engines, and, in the case of Claude, for improving code regarding numerical results.

Abbreviations

Abbreviation	Meaning
KS	Kochen–Specker
KSNC	Kochen–Specker noncontextuality
EWF	Extended Wigner's friend
FR	Frauchiger–Renner
LF	Local Friendliness
AOE	Absoluteness of Observed Events
LA	Local Agency
ROE	Relativity of Observed Events
NF	Noncontextual Friendliness
PBR	Pusey–Barrett–Rudolph
QFT	Quantum field theory
BH	Black hole
QRF	Quantum reference frame(s)

— 1 —

Introduction

"*What is proved by impossibility proofs is lack of imagination.*"
J.S. Bell, 1982 [1].

Two of the most prominent and persistent foundational puzzles in physics concern the measurement problem and the black hole information puzzle. These topics probe the limits of the two pillars of modern physics, quantum theory and general relativity. To date, there appear to be no fundamental constraints on the applicability of quantum theory, motivating the hypothesis of its universal validity. The measurement problem pushes the universality of quantum theory into the macroscopic world, or at least into regimes where a macroscopic notion of an observer exists, while black hole paradoxes do so into the regime of astrophysical gravitational phenomena. Recent results have sharpened the tension in the measurement problem by combining it with other distinctive features of quantum mechanics. In this thesis, we further develop and analyze these arguments in Part I. In Part II, we combine arguments of Part I with black hole physics, leading to new black hole paradoxes, highlight subtleties in black hole paradoxes more generally, and move towards combining quantum aspects with gravity.

In the remainder of this introduction, we provide some background for Parts I and II and present our main findings.

1.1 Part I: Universality in quantum foundations

1.1.1 The measurement problem, a universality hypothesis and Wigner's friend

The measurement problem [2–8], noted since the earliest days of quantum theory and one of the key outstanding challenges in quantum foundations [9], arises from two types of time evolution that can be prescribed to a quantum system: (i) a deterministic, unitary evolution of a closed quantum system, and (ii) an indeterministic wavefunction collapse after a system undergoes a measurement. But why do quantum systems in a superposition collapse to a single-valued outcome, or more precisely, to the corresponding eigenstate, upon being measured? What qualifies as an observer or a measurement, and which observers should apply evolution (ii) to a measured system, given that the system is ultimately part of a closed system governed by evolution (i)?

Besides Schrödinger's cat, another well-known illustration of this problem is the scenario of Wigner's friend [10]. In this thought experiment, involving Wigner and his friend, the friend measures a quantum system inside a sealed lab and claims to obtain a single-valued outcome. Wigner, remaining outside the sealed lab, considers his friend ultimately to be a quantum system as well, and describes the measurement as a unitary interaction of quantum systems. Consequently, Wigner describes his friend as being in a superposition of seeing different outcomes, at odds with the friend's claim. Although the descriptions of Wigner and his friend differ, there is no immediate contradiction concerning the actual value obtained by the friend. This is where recent extensions of the Wigner's friend scenario enter the discussion [5, 11–33], which derive contradictions concerning the actual outcome value obtained by the friend from seemingly innocuous assumptions. The technical novelty of these extensions is that they combine Wigner's friend with Bell nonlocality [34–36], a form of 'nonclassicality' of quantum theory, famously referred to as 'spooky action at a distance' by Einstein [37]. In doing so, they eliminate the counterfactuality required to derive a contradiction in usual Bell nonlocality scenarios.

The first so-called *extended Wigner's friend* argument, proposed by Brukner [5], laid the basis for the Frauchiger–Renner paradox [11] and the Local Friendliness no-go theorem [12, 13], two of the most celebrated extended Wigner's friend paradoxes, and the main topics of the first part of this thesis. The Frauchiger–Renner (FR) paradox involves several superobservers such as Wigner, who model their friends performing measurements inside sealed labs unitarily. The superobservers and friends reason about each other's

outcomes and derive a contradiction.

The Local Friendliness (LF) no-go theorem can be seen as a device-independent version of the Frauchiger–Renner paradox. Now, instead of reasoning about outcomes obtained by other agents[1], each superobserver can freely choose whether to ask their friend for their outcome or to perform other operations on the entire friend's lab. Assuming a notion of locality[2], i.e., that free choices can only affect outcome variables in their future, a contradiction is derived with the assumption that the measurement outcomes observed by the agents are absolute and single-valued, also referred to as Absoluteness of Observed Events[3].

The Frauchiger–Renner paradox and Local Friendliness no-go theorem can be argued to probe slightly different questions. The Local Friendliness no-go theorem shows that a weak notion of locality and Absoluteness of Observed Events are incompatible, assuming only the experimental validity of quantum theory in addition to the existence of superobservers. Arguments such as the Frauchiger–Renner paradox instead present us with the question of how agents can use quantum theory consistently, and, as we will argue, are particularly illuminating for going beyond the Absoluteness of Observed Events.

1.1.2 Part I: Contributions

Our main contributions in Part I of this thesis consist of a refined version of the Frauchiger–Renner paradox, connecting extended Wigner's friend scenarios to notions of nonclassicality, and an extended Wigner's friend argument based on the Pusey–Barrett–Rudolph theorem [38] that can be interpreted as an argument against the absoluteness of free choices. We now outline these results in more detail.

Refining the Frauchiger–Renner paradox. In Chapter 3, we present a refined version of the Frauchiger–Renner paradox [11]. In this paradox, several observers perform measurements in sealed labs, and are subsequently acted upon by their respective superobservers. The superobservers, having obtained their outcome, each make a prediction for the outcomes of other observers. After this, the superobservers meet up and share their predictions, and find that their predictions are contradictory. In contrast

[1] We will use the terms 'agent' and 'observer' interchangeably.

[2] This 'locality assumption' can also be seen as a causality requirement.

[3] Occasionally, to improve readability, we will refer to the Absoluteness of Observed Events simply as Absoluteness.

to the original Frauchiger–Renner paradox, a contradiction is obtained in each run of the protocol, and the required consistency condition for combining the reasoning of different agents is weakened. Notably, FR's consistency condition can be argued to be already violated in a classical paradox, the Sleeping Beauty problem, which we discuss in Chapter 6, as opposed to our weaker consistency condition. This weaker consistency requirement proves particularly interesting if one seeks to refute the Absoluteness of Observed Events; our refined FR paradox then suggests a principle, which we name the Relativity of Observed Events, to replace Absoluteness. Paraphrased as an extension of Peres' dictum [39], the principle of Relativity of Observed Events reads '*Unperformed experiments have no results, and unknown results have no values*.'

Connecting Wigner's friend with noncontextuality. In Chapter 4, we connect Wigner's friend arguments to noncontextuality [40–44]. Noncontextuality captures nonclassical features of quantum theory, i.e. how certain quantum phenomena admit no classical explanation. Two leading notions of noncontextuality are Kochen–Specker (KS) noncontextuality [40–42] and generalized noncontextuality [43, 44]. KS noncontextuality concerns the impossibility to assign outcome values to a set of measurements, independently of the *context* in which they take place. In other words, suppose we have measurements A, B, C, with experimental data from (i) context $\{A, B\}$ in which A is measured jointly with B, and (ii) context $\{A, C\}$ in which A is measured jointly with C; KS noncontextuality then requires that the outcome value assigned to A must be the same in each context. There exist correlations in quantum theory that fail this condition, and we say that quantum theory is *KS contextual*. Bell nonlocality can be seen as a special case of KS contextuality, with the joint measurements arising from an n-party structure. Generalized noncontextuality is defined within the framework of ontological models. Ontological models assign a space of ontic states to each system, with experimental preparations preparing a system in a distribution over ontic states, and measurements extracting information about that ontic state. Generalized noncontextuality then requires preparation procedures indistinguishable in any experiment, even when lacking full knowledge, to be represented identically in the ontological model, and similarly for transformations and measurement procedures.

To date, most extended Wigner's friend arguments are built on correlations that violate Bell nonlocality. In contrast, in this chapter we provide several new classes of extended Wigner's friend paradoxes employing noncontextual correlations, both Kochen–Specker [40–42] and generalized noncontextuality [43, 44]. Furthermore, we connect

conditions from the Local Friendliness assumptions on experimental correlations to conditions from Kochen–Specker noncontextuality.

An argument against the absoluteness of free choices. In Chapter 5, we provide an extended Wigner's friend argument based on the Pusey–Barrett–Rudolph (PBR) theorem [38]. The PBR theorem was originally presented as an argument against an epistemic interpretation of the quantum state, modulo some assumptions. We combine Wigner's friend and the scenario used by Pusey–Barrett–Rudolph to present a no-go theorem similar to the Local Friendliness no-go theorem, but as an argument against the absoluteness of free choices. Specifically, we use the same locality assumption of the Local Friendliness assumptions, but the relevant events in the assumption of Absoluteness of Observed Events, i.e. those that are unavailable after the protocol, concern free choices rather than measurement outcomes.

On Sleeping Beauty. Finally, in Chapter 6, we briefly discuss the Sleeping Beauty problem. Taking an operational viewpoint, we show how the tension in this problem arises from a difference in prior knowledge.

1.2 Part II: Universality and black holes

1.2.1 The black hole information problem and related paradoxes

Arguably, black holes rank among the most fascinating phenomena in physics, where intriguing puzzles arise at the intersection of quantum theory and gravity. In his seminal work, applying principles of quantum theory to black holes, Hawking showed that black holes radiate [45]. Moreover, he suggested this radiation to be purely thermal [46], which would lead to a loss of information as the black hole evaporates. Even if one had started from an initial pure state of collapsing matter, the radiation left after evaporation is purely thermal, an apparent nonunitary evolution. But where did this information go?

Other famous black hole puzzles involve objects and observers falling into the black hole, such as the black hole cloning argument [47–49] and the firewall paradox [49–53]. The black hole cloning paradox relies on the *no-cloning* feature of quantum theory [54–57], which states that, in quantum theory, one cannot clone unknown quantum states. That is, there cannot exist a quantum device which, given an arbitrary quantum state $|\psi\rangle$,

outputs two copies of this state, i.e., $|\psi\rangle \otimes |\psi\rangle$. In the black hole cloning paradox, one considers an observer Alice holding a quantum system S in an arbitrary quantum state $|\psi\rangle$, falling into the black hole. An outside observer Bob gathers all Hawking radiation coming from the black hole until late times. Under the standard assumptions of unitarity and that the information about the system S can be decoded from the Hawking radiation, Bob decodes the quantum state $|\psi\rangle$ from the Hawking radiation. But this leads to a violation of the no-cloning theorem, with the quantum state $|\psi\rangle$ being present both inside and outside the black hole.

The firewall paradox is derived from two main ingredients: (i) the monogamy of quantum entanglement [58–61], i.e. a system A cannot be highly entangled with both a system B and another disjoint system C, and (ii), the Page curve [62, 63]. The Page curve [62] gives the entanglement entropy over time of a 'typical' quantum subsystem R that grows in a global system consisting of two subsystems R and B (the radiation and black hole). Assuming a global pure state at all times, the entropy increases as R (the radiation) grows, reaches a maximum and then decreases again as purifying parts of B become part of R. We note that this analysis stems from a toy quantum model; in actual quantum gravity, identifying the radiation subsystem is in fact a subtle and important point [64–68]. The firewall argument then derives a contradiction as follows. Assuming the Page curve for the radiation [62], late-time radiation modes just outside the black hole should be highly entangled with the early radiation. However, according to quantum field theory on curved spacetime, which posits high entanglement across the horizon, these modes should also be highly entangled with modes inside the black hole, violating monogamy of entanglement.

While a full resolution of the black hole information puzzle remains elusive, much work has been done since, hinting that information is preserved. After some early arguments against Hawking's claim of information loss [69–73], a concrete argument in favor of unitarity came from AdS/CFT [49, 74]. More recently, the proposed existence of 'entanglement islands' inside the black hole yields a unitary Page curve for the radiation [68, 75–94], but whose physical interpretation remains heavily debated [65–68, 94–100]. Another interesting development is the identification of many extra degrees of freedom of asymptotically flat gravitating spacetimes, so-called *soft hair* [101–106], which could provide the missing piece Hawking was looking for when formulating his information-loss claim. Namely, soft hair endows a black hole spacetime with many more degrees of freedom for the radiation to entangle with, and may thus be a crucial factor

towards purification in a unitary evolution. Soft hair can be seen as arising from gauge transformations that act nontrivially on the boundary of spacetime [102]. Issues of gauge invariance, i.e. the nontrivial role played by apparent 'redundant' degrees of freedom, comprising diffeomorphism and coordinate invariance, may also be crucial in resolving the cloning and firewall paradoxes, as we briefly comment upon below and in the chapters referred to there.

1.2.2 Part II: Contributions

By combining the extended Wigner's friend arguments of Part I with black hole physics, we sharpen the cloning and firewall paradoxes, using similar assumptions. We criticize these assumptions by highlighting several important subtleties, which calls for more careful verification thereof using known techniques without committing to a specific theory of quantum gravity. We next present a calculation that uses some of these proposed methods in analyzing the detector response near a superposed black hole.

Wigner's friend's black hole adventure. In Chapter 8, we combine extended Wigner's friend arguments with black hole physics into a unified argument. This is an extension of the work by Hausmann and Renner [107], who drew analogies between Wigner's friend paradoxes and black hole puzzles. The cloning and firewall paradoxes raise sharp questions when combining the perspectives of different observers. However, they leave room for a more general (hypothetical) theory that could describe the different viewpoints of the observers, which could accommodate (unobservable) cloning or non-monogamy of entanglement, for example. The arguments of this chapter close this loophole, showing that no such theory exists if no observer can experimentally falsify the predictions of quantum theory.

In other words, we exhibit an inconsistency in the operational predictions prescribed by quantum theory under standard assumptions of black hole puzzles, such as the cloning and firewall paradoxes. After presenting our no-go results, we criticize these assumptions by highlighting important subtleties. Key elements in these subtleties include (i) quantum backreaction of the black hole leading to a superposed black hole, (ii) gauge invariance, i.e. the required invariance of true observables under gauge transformations, and (iii) observer-dependencies in describing the physics at hand. By emphasizing these issues, we invite a more careful verification of these assumptions towards resolving these puzzles. Among the methods that can be employed for this purpose are the use of quantum

reference frames and operational investigations of a black hole and its radiation. The next chapter illustrates a combination of these approaches in a simple, limited setting, without committing to a specific full theory of quantum gravity.

Detector near a superposed black hole. In Chapter 9, we analyze the response of a particle detector near a position-superposed BTZ black hole [108, 109] using aspects of quantum reference frames. Conservation laws [110, 111] suggest that an old black hole is highly superposed, for example with a highly superposed mass [112], also advocated for in Refs. [70, 113–117], and perhaps even more so when taking into account the vacuum degeneracy due to soft hair [102, 118]. This motivates the study of a superposed spacetime, and the particular choice of a position-superposed black hole as background (quantum) spacetime was motivated by the study of a detector's response near a mass-superposed black hole in literature by Foo et al. [119]. Our first result in this chapter is the derivation of a form of the Unruh–DeWitt interaction in a superposed spacetime using quantum reference frames. Secondly, Foo et al. [119] found peaks in the detector response, argued to be reminiscent of black hole mass quantization. By showing the absence of such peaks in a position-superposed black hole, this further strengthens the claim of such peaks being signatures of black hole mass quantization. Additionally, we derive analytical expressions for the spectrum probed by the detector, and show that, in the case of the mass-superposed black hole [119], the peaks in the detector response arise from singularities in the spectrum.

Conclusions and outlook. Finally, in Chapter 10, we close with some concluding remarks and future directions for both parts of this thesis.

1.3 Structure of thesis

The remainder of this thesis is organized as follows. Part I starts with Chapter 2, outlining preliminaries concerning Bell nonlocality, contextuality and extended Wigner's friend paradoxes, after which the main results of part I are presented in Chapters 3 to 6. Part II opens with some background on black hole paradoxes in Chapter 7, after which we present the main results of part II in Chapters 8 and 9. Finally, we conclude in Chapter 10.

This thesis is organized so that each chapter can be read individually, and we avoid extensive cross-referencing between chapters. Therefore, relevant assumptions are

repeated in each chapter, occasionally with slightly different formulations to suit the specific result and purpose.

Part I

Universality in quantum foundations

The first part of this thesis concerns foundational puzzles in quantum mechanics, more specifically, extensions of the Wigner's friend scenario, closely related to the measurement problem. We will analyze, refine and present several novel extended Wigner's friend arguments, and examine a related classical paradox. After the preliminaries in Chapter 2, we present the main results of this part.

In Chapter 3 we refine the Frauchiger–Renner paradox, constructing a scenario that relies on the strongly nonlocal GHZ–Mermin scenario. Whereas most existing extended Wigner's friend arguments rely on nonlocal correlations, we present several extended Wigner's friend arguments based on noncontextuality in Chapter 4. Additionally, we investigate the link between Local Friendliness and noncontextuality inequalities. Next, in Chapter 5 we combine Wigner's friend with the Pusey–Barrett–Rudolph theorem, which was originally constructed to question the validity of epistemic interpretations of the quantum state. The resulting paradox questions the absoluteness of free choices, in the same way that extended Wigner's friend arguments such as Local Friendliness question the absoluteness of outcomes of performed measurements under a locality assumption. Finally, in Chapter 6 we present a discussion of the Sleeping Beauty problem, a classical paradox in probability theory and philosophy that bears similarities to Wigner's friend arguments, especially to the Frauchiger–Renner paradox. A key element of this analysis of the Sleeping Beauty problem is a novel operational viewpoint on the questions asked.

The results of Chapters 3, 4 and 6 are based on Refs. [19], [20–22] and [120], respectively. The results of Chapter 5 are based on Ref. [121].

— 2 —

Quantum preliminaries

In this section we outline the preliminaries to our work of Part I, which comprises extensions of Wigner's thought experiment. We start with outlining the conventions we use regarding the formalism of quantum theory, after which we move to notions of nonclassicality, which help us to understand how quantum theory differs from classical theory, and how this quantum advantage could be harnessed. Next, we move to the measurement problem and Wigner's original thought experiment. We end by discussing some recent extended Wigner's friend arguments in literature.

Sections 2.3.1 and 2.4, Section 2.5 and Section 2.2.2 contain parts copied from Ref. [19, Section 2], Ref. [21, Section 2.2] and Ref. [22, Section 3], respectively.

2.1 A short recap of quantum theory

2.1.1 Notation in linear algebra

Quantum theory associates systems with states, which can be composed, and upon which transformations can be performed and information can be extracted by means of measurements. For our introduction to quantum theory in this chapter, we restrict ourselves first to finite-dimensional systems, unless stated otherwise.

We denote the inner product of the Hilbert space by $\langle\cdot|\cdot\rangle$, and the set of homomorphisms between two vector spaces A, B, i.e. the linear maps between A and B, by $\mathrm{Hom}(A, B)$. A homomorphism from a vector space to itself is called an endomorphism. An endomorphism on a Hilbert space $\mathcal{H}$ is also called an operator on $\mathcal{H}$. In the bra-ket notation, a vector $\alpha \in \mathcal{H}$ is associated with a *ket* $|\alpha\rangle$, which can also be seen as an

element in $\mathrm{Hom}(\mathbb{C}, \mathcal{H})$ as $|\alpha\rangle : \gamma \mapsto \alpha\gamma$. The adjoint $|\alpha\rangle^*$ of this mapping is called a *bra*, and is an element of the dual space $\mathcal{H}^* = \mathrm{Hom}(\mathcal{H}, \mathbb{C})$ acting as $\langle\alpha| : \beta \mapsto \langle\alpha|\beta\rangle$. It follows that $\langle\alpha|\ |\beta\rangle = \langle\alpha|\beta\rangle$. The composition $|\beta\rangle\langle\alpha|$ is an operator on $\mathcal{H}$, and any operator can be written as a linear combination of such elements. The tensor product is denoted by $\otimes$, the trace by Tr and the partial trace over a subsystem B by Tr_B.

2.1.2 Quantum theory postulates

The postulates of finite-dimensional[1] quantum theory can be summarized as follows. We refer the reader to Refs. [57, 128, 129] for more details.

a. State space. The state of a quantum system (created through some preparation procedure) is associated with a *density operator* ρ on a complex Hilbert space, that is, a normalized positive semi-definite operator on $\mathcal{H}$, i.e. $\rho \geq 0$ and $\mathrm{Tr}\,\rho = 1$. The set of all density operators associated with a system is also referred to as the state space denoted by $S(\mathcal{H})$. A density operator is called *pure* if it can be written as $\rho = |\psi\rangle\langle\psi|$. A system prepared probabilistically, i.e. with probability p_i in state ρ_i, corresponds to the density operator $\sum_i p_i\rho_i$.

b. Composing quantum systems. The composition of systems is facilitated by the tensor product $\otimes$. In a composite quantum system $A \otimes B$ with Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, the system A with Hilbert space $\mathcal{H}_A$ is called a subsystem of $A \otimes B$. Given a state ρ_{AB} for $A \otimes B$, the reduced state on system A is given by the partial trace over B, $\rho_A = \mathrm{Tr}_B[\rho_{AB}]$. Given the composite system $A \otimes B$, not all states are *separable*, i.e. there exist *entangled* states ρ_{AB} which cannot be written as $\sum_i p_i \rho_A^{(i)} \otimes \rho_B^{(i)}$.

c. Time evolution. The time evolution of a closed quantum (sub)system corresponds to a unitary U on the state space of that system, i.e. $\rho \mapsto U\rho U^\dagger$.

d. Measurements. A projective measurement on a quantum (sub)system A is associated with an *observable* O_A, i.e. a Hermitian operator, decomposed into projectors as $O_A = \sum_{x\in X} xP_x$ with outcomes $x \in X$; P_x is the projector onto the eigenspace with eigenvalue x. The probability of observing outcome

[1] For infinite-dimensional quantum theory and in the case of infinitely many degrees of freedom (in local quantum field theories, for instance), we refer the reader to Refs. [122–124] and Refs. [125, 126], respectively; see also the paragraph *Towards quantum field theory and beyond* in Section 2.1.3 below. For a measurement scheme with state update rules in quantum field theory, see also Ref. [127].

x when measuring O_A on a density matrix ρ_{AB} is determined by the Born rule:

$$p(x|O_A, \rho_{AB}) = \mathrm{Tr}[P_x \otimes \mathbb{1}_B \, \rho_{AB}]. \tag{2.1}$$

The post-measurement state conditioned on this outcome is then given by

$$\rho_{AB|x} = \frac{1}{p(x|O_A, \rho_{AB})} (P_x \otimes \mathbb{1}_B) \, \rho_{AB} \, (P_x \otimes \mathbb{1}_B). \tag{2.2}$$

This state update is also referred to as the *wavefunction collapse*. More general measurements are introduced below.

2.1.3 Remarks

We now outline some remarks about the above postulates. For the first five remarks concerning measurements and time evolution we refer to Refs. [57, 128, 129] for more details.

Open vs. closed systems. The above postulates are presented in their traditional (restricted) form. In modern quantum information theory, states are represented by density operators and all physical processes (including measurements) are described by trace-preserving completely positive maps (TPCPMs). Closed systems evolve unitarily, while open systems evolve according to TPCPMs. Further, the above measurement postulate concerns only sharp (or ideal or projective) measurements, while the most general form of a measurement is determined by a quantum instrument. We briefly review these generalizations in the following four remarks.

Generalizing time evolution. The time evolution and measurement postulates above refer to isolated systems, i.e. when no other involved systems were discarded (traced over). More generally, the evolution on a subsystem is given by a trace-preserving completely positive map $\mathcal{E}$, where completely positive means that for any Hilbert space $\mathcal{H}_R$ the map $\mathcal{E} \otimes \mathcal{I}_R$ is positive. For any TPCPM on $\mathcal{H}_A$ (acting on density operators on $\mathcal{H}_A$), there always exists a Hilbert space $\mathcal{H}_E$ and a unitary U on $\mathcal{H}_A \otimes \mathcal{H}_E$ such that $\mathcal{E}$ is the restriction of U to subsystem A (Stinespring dilation theorem): $E(\rho) = \mathrm{Tr}_E \left[U \, (\rho \otimes \sigma_E) \, U^\dagger \right]$.

Operator-sum representation. We can write any TPCPM $\mathcal{E} \in \mathrm{Hom}(\mathcal{H}_A, \mathcal{H}_B)$ in the operator-sum representation as

$$\mathcal{E} : S_A \mapsto \sum_x M_x S_A M_x^\dagger \tag{2.3}$$

with $M_x \in \mathrm{Hom}(\mathcal{H}_A, \mathcal{H}_B)$ and $\sum_x M_x^\dagger M_x = \mathbb{1}_A$; $\{M_x\}$ are called the Kraus operators. This association of $\{M_x\}_x$ to $\mathcal{E}$ is not unique. Any mapping of this form is also a TPCPM.

Measurements as unitaries. Any projective measurement $O_A = \sum_x xP_x$ on ρ_{AB} can be seen as a TPCPM, and as a result, also as a unitary. One constructs a Hilbert space $\mathcal{H}_X$ with orthonormal basis $\{|e_x\rangle\}_x$ corresponding to the outcomes $x \in X$. The post-measurement state is then given by

$$\rho'_{XAB} = \sum_x p(x) \left|e_x\right\rangle \left\langle e_x\right| \otimes \rho_{AB|x} = \sum_x M_x \rho M_x^\dagger, \tag{2.4}$$

a TPCPM, with $p(x) = p(x|O_A, \rho_{AB})$ and $M_x = |e_x\rangle \otimes P_x \otimes \mathbb{1}_B$ satisfying $\sum_x M_x^\dagger M_x = \mathbb{1}_{AB}$.

Generalizing measurements. Regarding measurements, we can allow more generally a measurement on $\rho \in S(\mathcal{H}_A)$ to be associated with a set of operators $\{M_x\}_x$ on $\mathcal{H}_A$ satisfying $\sum_x M_x^\dagger M_x = \mathbb{1}$. Namely, letting $\mathcal{H}_X$ be a Hilbert space of dimension equal to the size of the outcome set with orthonormal basis $\{|e_x\rangle\}_x$, we can apply the isometry $V = \sum_x |e_x\rangle \otimes M_x$ acting on a state $\rho_A \in S(\mathcal{H}_A)$, and next applying the projective measurement $\{|e_x\rangle \langle e_x| \otimes \mathbb{1}_A\}_x$ on $V\rho V^\dagger$. The resulting probabilities are given by $p(x) = \mathrm{Tr}[M_x \rho_A M_x^\dagger]$, and the post-measurement state by

$$\rho'_{A|x} = \frac{1}{p(x)} M_x \rho M_x^\dagger. \tag{2.5}$$

Note that the probabilities are in fact determined by $N_x = M_x^\dagger M_x$ satisfying $\sum_x N_x = \mathbb{1}$ by

$$p(x) = \mathrm{Tr}[N_x \rho]. \tag{2.6}$$

Such positive Hermitian operators $\{N_x\}_x$ satisfying $\sum_x N_x = \mathbb{1}$ are called a positive operator-valued measure (POVM), and thus fully characterize the measurement probabilities. With any such POVM N_x one can associate a CPM describing the measurement, for example by $M_x = \sqrt{N_x}$. This association is not unique, and as a result a POVM alone cannot be used to infer the post-measurement state; the latter depends on the Kraus operators or, equivalently, the chosen quantum instrument.

A quantum instrument, the most general description of a measurement that determines the post-measurement state, is a set $\{\mathcal{E}_x\}$ of completely positive trace-nonincreasing maps such that $\mathcal{E} = \sum_x \mathcal{E}_x$ is trace-preserving. The outcome probabilities are given by $p(x) = \mathrm{Tr}[\mathcal{E}_x \rho_A]$, and the post-measurement state by

$$\rho_{A|x} = \frac{1}{p(x)} \mathcal{E}_x(\rho_A). \tag{2.7}$$

Each CPM $\mathcal{E}_x$ is associated with a set of Kraus operators by $\mathcal{E}_x(\rho_A) = \sum_k M_{x,k}\rho_A M_{x,k}^\dagger$ and POVM elements $N_x = \sum_k M_{x,k}^\dagger M_{x,k}$. The sum over the index k in the Kraus representation of each element of an instrument can intuitively be understood as summing over unobserved degrees of freedom (for instance arising from tracing out an environment), or as coarse-graining over a more fine-grained set of measurement outcomes [129]. It follows from the above discussion that a quantum instrument can be implemented by a projective measurement on an extended system, and may thus be modeled unitarily as in the paragraph *Measurements as unitaries* above.

Pure states. Above, we have expressed a pure state as a density operator ρ of the form $|\psi\rangle\langle\psi|$; an equivalent condition is that $\mathrm{Tr}[\rho^2] = 1$. Often, a pure state is expressed simply as the normalized vector $|\psi\rangle$ in the Hilbert space. In this case, there is a phase ambiguity, in that $e^{i\theta}|\psi\rangle$ and $|\psi\rangle$ correspond to the same physical state, when considered isolated from everything else. This 'redundant' phase degree of freedom leads to representations of (classical) symmetry groups on the Hilbert space being projective representations in quantum theory, and may lead to a type of 'quantum anomaly', such as the anomalous central charge in a conformal quantum field theory for example [130]. Any mixed state can be purified, i.e. for any state ρ_A there exists a system B and a pure state ρ_{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$ such that $\mathrm{Tr}_B\, \rho_{AB} = \rho_A$.

The qubit. A qubit is the quantum version of a classical bit, i.e. a system with a two-dimensional Hilbert space. A classical bit can be in two states, $|0\rangle$ and $|1\rangle$. These states also form a basis for the two-dimensional Hilbert space of a qubit, but now the qubit can be in a superposition of $|0\rangle$ and $|1\rangle$, i.e. a linear combination of these basis states,

$$\alpha|0\rangle + \beta|1\rangle\,, \tag{2.8}$$

with α, β complex numbers, normalized such that $|\alpha|^2 + |\beta|^2 = 1$.

Relating to the Hamiltonian. In continuous-time systems, the time evolution postulate is sometimes phrased in terms of the Hamiltonian, a Hermitian operator whose eigenvalues are the energy levels of a system, reading as

$$i\hbar\partial_t|\psi\rangle = H|\psi\rangle\,. \tag{2.9}$$

In fact, unitary time evolution can be seen as arising from the interaction of a system with a clock, as in the Wheeler–DeWitt equation [131–147], with the Hamiltonian generating

the unitary representation [148]. In fact, doing the same for spatial coordinates leads to a quantum version of Einstein's rods and clocks [147, 149–152], closely related to the paradigm of quantum reference frames that we will encounter in Part II of this thesis.

Alternative postulates. Above, we have adopted a standard way of expressing the postulates of quantum mechanics. However, there are alternative ways of axiomatizing or deriving quantum theory. In fact, an active area of research in quantum foundations concerns reconstructions of quantum theory and derivations from some core physical, informational or structural principles, aiming to answer the questions such as 'What is quantum theory?' and 'Why quantum theory?'. Developments in this field (including efforts reducing the number of postulates or towards a full derivation) can be categorized into operational derivations [153–164], which posit physically motivated axioms on the structure of states, transformations and measurements; derivations that aim to derive quantum theory from information-theoretic principles on experimental correlations [165–175]; agent-centric derivations [176–180]; and diagrammatic reconstructions [181] inspired by categorical quantum mechanics [182–184].

Towards quantum field theory and beyond. Moving from finite-dimensional quantum theory to local quantum field theory (QFT), i.e. an infinite-dimensional quantum theory with infinitely many degrees of freedom $\phi(x,t)$ (formally) labeled by spacetime points, comes with many difficulties. The notions of subsystems, factorizations and tracing over subsystems must be treated with care, and instead of functions and operators, one works with fields as operator-valued distributions instead. As a result, divergences in the calculation of certain quantities arise, and are typically handled by regularization and renormalization procedures, or by the introduction of novel concepts. In the case of special-relativistic QFT, or a QFT with some other symmetry group, one must incorporate Lorentz invariance as well, and notions such as particles become observer-dependent. In fact, Lorentz and Poincaré invariance, in the form of relativistic wave equations [185–188] or Poincaré representations [189], and locality can be argued to provide the conceptual leap from quantum mechanics to quantum field theory [130]. Conceptually, in light of the use of methods such as regularization and renormalization, one can view our use of (current) quantum field theories as effective theories [130, 190–192].

In curved spacetime, one has the additional difficulty of gauge symmetries associated with the curved spacetime background, namely diffeomorphism invariance and local Lorentz invariance in the presence of spin structures, the multitude of time parameters

and the potential absence of a preferred one to define positive frequency modes and build Fock spaces, and (quantum) backreaction of the background structure including the metric— perhaps unavoidable crucial features rather than problems. Moreover, one must rule out some states on a physical basis, such as the Hadamard condition [126, 193–196]. Such Hadamard states have a short-distance singularity structure matching that of states in Minkowski spacetime with finite energy density, in line with Einstein's equivalence principle, and is required for example to avoid infinite fluctuations of the quantum field [196].

These many problems lead to a multitude of approaches to quantum field theory (on curved spacetime), such as canonical or Fock space quantization [126, 130, 197], algebraic quantum field theory [125, 127, 198, 199], perturbative algebraic quantum field theory [195, 200, 201], path integral methods [202–204], deformation quantization [205, 206], with several possibilities of dealing with gauge symmetries [207–213]. Another recent approach to quantum field theory, still under active development, employs non-standard analysis formulated in the framework of categorical quantum mechanics [214, 215], allowing for QFT postulates closely resembling those of standard quantum theory.

The approach we will use in Part II is the canonical quantization approach to quantum field theory. We will not present an outline of this approach, but refer the reader to the literature [126, 130, 197].

2.2 Notions of nonclassicality

Having outlined the foundational axioms of quantum theory, we move to the study of nonclassicality in quantum theory. The motivation for studying nonclassicality is twofold. A first motivation is practical, in that it facilitates the harnessing of quantum advantage [216–221], i.e., how the use of quantum theory's resources enables improved performance in certain tasks. Secondly, studying quantum theory's nonclassicality deepens our understanding of quantum theory, how it defies our classical intuition, complementary to the study of reconstructions and derivations of quantum theory outlined in Section 2.1.3.

Why would nature, or at least a big part of it, be described by quantum theory? And how does quantum theory differ from classical theory? Is it possible that underlying quantum theory there is still some classical reality, potentially inaccessible to us, that explains these mysterious phenomena in nature? As Feynman famously expressed: 'I

think I can safely say that nobody understands quantum mechanics' [222], referring to its counterintuitive properties.

Notions of noncontextuality. One such example defying our classical intuition is the non-existence of a common cause that classically and fully explains the correlations between distant quantum systems, without hidden causal connections, formalized as Bell nonlocality [34, 35, 223–225]. Bell nonlocality can be seen as a special instance of Kochen–Specker (KS) contextuality [41], which captures the nonexistence of hidden classical variables that predetermine the outcomes of experiments. Alternatively, also KS noncontextuality can be seen as capturing the idea of a classical causal model explaining empirical data, under the additional assumption that an operational symmetry is preserved in the classical model [224, 226]. Another leading notion of nonclassicality is generalized contextuality, which applies to preparations, transformations and measurements. KS noncontextuality can be seen as an instance of generalized noncontextuality under the additional assumption of outcome determinism for the measurements [227].

No fine-tuning. Central to notions of nonclassicality are the existence of some 'classical model' representing the experiment to explain the correlations, together with notions of no fine-tuning [223, 224, 226, 228], with causal no fine-tuning for Bell locality and Kochen–Specker noncontextuality [223, 224, 226] and operational no fine-tuning for generalized noncontextuality [228]. No fine-tuning of a classical model expresses the idea that the model variables are nonconspiratorial, i.e. that the underlying classical model follows the causal structure of the experiment, or reflects properties of the operational theory at hand. In defining generalized noncontextuality, if two preparations, measurements or transformations are empirically indistinguishable, operational no fine-tuning requires that they are represented the same in the ontological models framework of classical realism, also known as Leibniz's principle of the identity of empirically indiscernibles [229]. Operational fine-tuning [228] also captures the time-symmetry fine-tuning that a classical realism explanation of quantum theory requires [230]. Causal no fine-tuning or faithfulness of a causal model means that there are no additional conditional independence relations that are not implied by the causal diagram; if this were not the case, slightly varying parameters of the causal model generically leads to a violation of these additional conditional independences. Bell nonlocality and KS contextuality demonstrate that quantum theory does not admit of a causal no fine-tuned classical model [223, 224, 226], and generalized noncontextuality shows that quantum theory does not admit of an operational no-fine tuned classical (realist) model. Taken together, quantum theory does not admit of a classical realist explanation free of fine-tunings in either sense.

Beyond realism? What these no-go theorems expressing the nonclassicality of quantum theory tell us is that, when adhering to the principle of no fine-tuning, classical, realism models cannot explain the experimental data of nature. Moreover, within the ontological models framework of classical realism, one can also ask whether the quantum wavefunction should be interpreted ontically or epistemically; does it capture the underlying reality or merely a lack of knowledge of a system? The PBR theorem [38, 231] demonstrates that the wavefunction cannot be epistemic, modulo some (relatively weak) assumptions. However, these classical models are *realist* representations of the world, assuming the existence of 'states of reality', independently of any observer, independently of how or by whom they are accessed, using a wording similar to Ref. [5]. A possible resolution for a deeper understanding of quantum theory thus lies in allowing for relational descriptions and observer-dependencies. This resolution can be argued to be in line with a resolution to Wigner's friend paradoxes that we will explore in Chapter 3, the Relativity of Observed Events, expressing that certain facts may be relative.

2.2.1 Kochen–Specker contextuality and Bell nonlocality

This section follows Section 2 of Ref. [21], nearly verbatim. We present a minimal amount of mathematics required; for a comprehensive introduction, we refer the reader to Refs. [41, 42, 232].

Kochen–Specker noncontextuality. Let $\mathcal{M}$ be a finite set of measurements, of which some subsets are compatible, that is, jointly measurable [233, 234]. These subsets are called *contexts*, and a given measurement may appear in more than one context. A context is said to be maximal in a scenario if it is not a subset of any other context. For example, if we have 3 measurements labeled by M_1, M_2, M_3 in the same context (denoted as a set $\{M_1, M_2, M_3\}$), we can jointly measure them, obtaining a joint outcome (m_1, m_2, m_3). A *measurement scenario* $(\mathcal{M}, \mathcal{C}, O)$ is defined by a finite set of measurements $\mathcal{M}$, a finite set of maximal contexts $\mathcal{C}$, and a set of possible outcomes O for each measurement.

One can represent such a measurement scenario by a compatibility hypergraph—a hypergraph that associates a measurement to each node of the graph and a hyperedge to each context [232]. In specific measurement scenarios, it is sufficient to consider only the 2-skeleton of the compatibility hypergraph, known as the *compatibility graph* [232, Def. 2.2, pg. 15] (see also Ref. [235]), in which case contexts are denoted by cliques in the graph, see Figure 2.1 for an example of this.

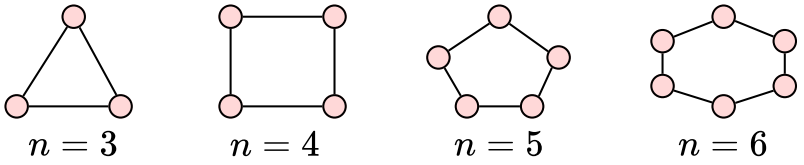

Figure 2.1: n-**cycle compatibility hypergraphs.** Cycle graphs C_n for $n = 3, 4, 5, 6$. Each graph defines a compatibility structure that can be associated with measurement scenarios; each node is associated with a measurement and the measurements connected by an edge are compatible and form a context. Graph C_4 depicts the compatibility structure of the bipartite Bell scenarios leading to the Clauser–Horne–Shimony–Holt (CHSH) inequality [236]. Graph C_5 depicts the compatibility structure leading to the Klyachko–Can–Binicioğlu–Shumovsky (KCBS) inequality [237].

A behavior in such a measurement scenario is a collection of *empirical correlations*, one for each context C, over the outcomes for the measurements in that context. An empirical correlation is said to satisfy *no-disturbance* if, for any measurement appearing in more than one context, the marginal probabilities over its outcomes are independent of context. For example, if M_1 is in the context with $C_1 \coloneqq \{M_1, M_2\}$ and the context $C_2 \coloneqq \{M_1, M_3\}$, namely, M_1 is compatible with M_2 and M_3, then the empirical correlation $p_{C_1}(m_1, m_2)$ of the first context and the empirical correlation $p_{C_2}(m_1, m_3)$ of the second context agree on their overlap[2], namely the probability for m_1:

$$\sum_{m_2} p_{C_1}(m_1, m_2) = \sum_{m_3} p_{C_2}(m_1, m_3). \tag{2.10}$$

A behavior is said to be Kochen–Specker noncontextual (KSNC) if there is a single 'global' probability distribution $p_{NC}(\cdot)$ over all measurements in $\mathcal{M}$ whose marginals reproduce the empirical correlations for each context. Note that any KSNC behavior automatically satisfies no-disturbance, so it only makes sense to consider KSNC for no-disturbing behaviors. The existence of a global distribution reproducing all empirical correlations is equivalent to the existence of a *deterministic hidden variable model*, or equivalently, a *factorizable* hidden variable model. Such deterministic hidden variable model is also referred to as an outcome-deterministic ontological model, as we define more carefully in Section 2.2.2. This means that there exists a hidden variable, or ontic state λ, potentially sampled over in each round of the experiment, that deterministically assigns an outcome to each possible measurement. If for some particular compatibility hypergraph, there is a behavior (or correlations) with no such joint probability distribution, then one calls such a behavior KS contextual.

[2]We say that distributions *agree on their overlap* if they have the same marginals on the intersection of their domains.

The set of all empirical correlations satisfying these assumptions in a given scenario is known as the Kochen–Specker noncontextuality polytope and is a function of the compatibility hypergraph for the scenario [42, 238–244]. The set of behaviors satisfying KS noncontextuality is characterized by *KS noncontextuality inequalities* [41, 237, 239, 245, 246].

Logical and strong contextuality. Contextual correlations are called *logically contextual* (or possibilistically contextual) if the proof that no joint probability distribution exists can be given while only keeping track of whether or not a given combination of measurement outcomes is possible or not, i.e., by looking at the support of the probability distributions in the argument [42, 238]. The contextuality is said to be *strong* if moreover no single outcome assignment within a context can be extended to a global assignment compatible with the given behavior.

Example: the KCBS scenario [237]. Consider five binary-outcome measurements $\{M_i\}_{i\in\{1,2,\ldots,5\}}$, where the pairs $\{(1,2),(2,3),(3,4),(4,5),(5,1)\}$ are jointly measurable, with compatibility graph C_5 as in Fig. 2.1. Imagine that the system is prepared such that the observations for these five joint measurements satisfy

$$p(1,1|1,2) = 0, \tag{2.11a}$$
$$p(0,0|2,3) = 0, \tag{2.11b}$$
$$p(1,1|3,4) = 0, \tag{2.11c}$$
$$p(0,0|4,5) = 0, \tag{2.11d}$$

and

$$p(0,1|5,1) \neq 0. \tag{2.12}$$

Denoting the deterministic assignment of the outcome of measurement M_i as m_i, Eq. (2.11) implies that

$$m_1 = 1 \Rightarrow m_2 = 0 \Rightarrow m_3 = 1 \Rightarrow m_4 = 0 \Rightarrow m_5 = 1. \tag{2.13}$$

Therefore, in every run of the experiment where the outcome of M_1 is 1, the outcome of M_5 must be 0. However, Eq. (2.12) tells us that in some runs of the protocol, the outcome of M_5 is 0 and the outcome of M_1 is 1. Thus, there is no Kochen–Specker noncontextual assignment for these runs. A quantum realization of this contextual behavior is given in Section 4.1.2.

Bell nonlocality. A special case of a measurement scenario is a Bell scenario, in which the measurement contexts arise from a party structure where each party can choose from a certain set of measurements to perform on the system they own. An example is the two-party Bell scenario. Consider Alice and Bob, choosing measurement settings x and y, obtaining outcomes a and b, respectively. The behavior is then given by the probabilities $p(a, b|x, y)$. The condition of no-disturbance is also referred to as *no-signaling* in the case of a Bell scenario (with terminology derived from no-superluminal signaling), in this example given by $p(a|x, y) = p(a|x)$ and $p(b|x, y) = p(b|y)$. The requirement of the existence of a global distribution for this behavior, or equivalently [42, 247], a factorizable hidden variable model, corresponds to the requirement of Bell's local causality:

$$p(a, b|x, y) = \sum_{\lambda} p(\lambda)p(a|x, \lambda)p(b|y, \lambda) \tag{2.14}$$

with $\lambda \in \Lambda$ the hidden variables, or ontic states, of the system. In this case, the Kochen–Specker noncontextuality polytope thus coincides with the Bell polytope [42, 238–244]—that is, the polytope of correlations consistent with Bell's local causality.

Compatibility, commutativity and no-disturbance for (un)sharp measurements. A finite set of measurements is said to be jointly measurable, or compatible, if they can all be obtained from coarse-grainings of a single measurement [233, 234, 248, 249]. In quantum theory, sharp or ideal measurements correspond to projective measurements. More generally, a theory-independent definition of a sharp (or ideal) measurement is as follows [175, 235, 250]: a measurement is sharp (or ideal) if (i) it yields the same outcome when performed consecutive times, (ii) it only disturbs measurements that are incompatible with it, and (iii) all its coarse grainings have realizations satisfying (i) and (ii).

In quantum theory, the observables of two compatible projective measurements commute. The state update rule with projectors implies that two compatible projective measurements, that is, whose observables commute, are non-disturbing; applying the sequence of measurements ABA for projective measurements A, B implies that the second A-measurement yields the same outcome as the first. In fact, for projective measurements, the notion of commutativity, compatibility and no-disturbance are equivalent [41, 248]. This is not true anymore for generalized, non-projective measurements, also referred to as unsharp or nonideal measurements; for non-projective measurements such as POVMs, the notions of commutativity, joint measurability and no-disturbance are no longer equivalent [41]. For example, for POVMs, pairwise compatibility of a set of

measurements does not imply global compatibility. Pairwise commutativity of a set of POVMs does imply global compatibility, but the converse is not true [234].

Outcome-determinism and unsharp measurements.[3] As we have seen, requiring KSNC can also be seen as requiring the existence of an outcome-deterministic ontological model. It has been argued that KSNC, containing this assumption of outcome-determinism, cannot be applied to both sharp and unsharp measurements alike [227]. One can motivate the assumption of outcome determinism for sharp quantum[4] measurements (and also drop the assumption of outcome determinism for unsharp measurements) by appealing to the notion of generalized noncontextuality presented in Section 2.2.2 [43, 227, 251].

However, regarding the question of when outcome-determinism is justified, we have seen that Bell nonlocality is an instance of KSNC, and in the case of Bell nonlocality, outcome-determinism is justified as an instance of Bell's local causality assumption, or as a causal no fine-tuning in a classical causal model [223, 224, 226]. Moreover, KSNC can be derived similarly under a causal no fine-tuning argument together with a symmetry assumption [224, 226]. So, where could this story go wrong for unsharp measurements?

One potential viewpoint on this matter is the following, with a crucial role for no-disturbance. It could very well be that the description of POVMs is not enough to apply the causal no fine-tuning argument to argue for Bell locality and KS noncontextuality as plausible notions of classicality, as it does not capture the complete description of the experimental set-up. Namely, there exist many projective measurements on a larger system that give the same POVM. In particular, if one has two no-disturbing POVMs, it could be that in the larger system (including an ancilla and environment, for example) that comprises the experimental set-up described with PVMs, these additional systems introduce correlations leading to the PVMs violating the no-disturbance condition. If, on the other hand, in the full experimental set-up, the PVMs of which the POVMs are derived still satisfy the no-disturbance condition, then the causal no fine-tuning conditions employing the no-disturbance condition can be argued to be justified. The crucial role for no-disturbance means that this argument is still compatible with the criticism of applying outcome determinism to just any POVM [227].

This is in line with how literature deals with the case of unsharp measurements and

[3] I thank Raman Choudhary for many interesting discussions on questions related to sharpness, outcome determinism and noncontextuality.

[4] Generalizing this proof of outcome-determinism from PVMs to theory-independent sharp measurements requires properties of perfect predictability of sharp measurements with eigenstates and operational indistinguishability for sharp measurements when mixing over all their respective eigenstates [227].

experimental imperfections. One approach, on a case-by-case basis, is to pose assumptions on the noise in the hidden-variable model, while typically performing additional measurements to aid this characterization and potentially modifying the noncontextuality inequalities [41]. Another is by devising noise-robust inequalities for KS scenarios within the framework of generalized noncontextuality [249, 252, 253].

2.2.2 Generalized noncontextuality

Another notion of noncontextuality, referred to as generalized noncontextuality, was introduced by Spekkens in Ref. [43], and further developed in Refs. [44, 227, 249, 254–258]. An operational theory consists of preparations, measurements and transformations. An ontological model for an operational theory is a classical realist description of a physical system as mediating between the preparation procedure and the measurement. Inspired by Leibniz principle of the ontological identity of empirical indiscernibles [229], the idea of generalized noncontextuality is that all operations, transformations and measurements that are statistically indistinguishable should have the same ontological representation.

An operational theory. An *operational theory* is a quadruple $T = (\mathcal{P}, \mathcal{T}, \mathcal{M}, p)$, where $\mathcal{P}, \mathcal{T}, \mathcal{M}$ are sets of preparation, transformation and measurement procedures. Furthermore, for each preparation P, transformation T and measurement M, there is a probability distribution $p(k|M, T, P)$ over the outcomes of the measurement M on the T-transformed preparation P. The underlying physical intuition is that a system is prepared in the lab using procedure P, transformed under T and measured in measurement M, giving outcome k with probability $p(k|M, P)$. Notably, also unsharp measurements, i.e. POVMs in quantum theory, are also allowed. We also refer to the combination (k, M) of a measurement M and outcome k as the effect $k|M$ or $[k|M]$.

Probabilistic mixtures of preparations are considered again a preparation, denoted by $\sum_i p_i P_i$ for performing preparation P_i with probability p_i. Statistically indistinguishable preparations are called operationally equivalent, and similarly for measurements and transformations, as defined below in *Operational equivalences and generalized noncontextuality*.

Ontological models. The ontological models framework [259] formally provides a realist explanation of the statistics of a physical theory by postulating the existence of physical systems with properties (ontic states) represented by points λ in a measurable

set Λ (the ontic state space). The statistics predicted by the theory are reproduced, in the ontological model, via the law of classical total probability:

$$p(k|M,P) = \sum_{\lambda} p(k|M,\lambda)p(\lambda|P), \tag{2.15}$$

where the preparation P is represented by a probability distribution $p(\lambda|P)$, and the measurement M with outcome k is represented by a probability distribution $p(k|M,\lambda)$, also referred to as the response or indicator function. A transformation is represented by a transition matrix $p(\lambda|\lambda',T)$, giving the probability density for a transition from the ontic state λ' to the ontic state λ.

The distributions $p(\lambda|P), p(k|M,\lambda), p(\lambda|\lambda',T)$ associated with a preparation procedure, measurement and transformations are often denoted by $\mu(\lambda|P), \xi(k|M,\lambda), \Gamma_T(\lambda|\lambda')$.

Operational equivalences and introducing generalized noncontextuality. Generalized noncontextuality requires that operational equivalences predicted by a physical theory correspond to identities in the ontological model of the theory. For instance, if two preparations P and P' are operationally equivalent, denoted $P \sim P'$, if they cannot be distinguished by any measurement, i.e. if for all possible measurements M with outcomes k, $p(k|M,P) = p(k|M,P')$. An ontological model is then said to satisfy *preparation noncontextuality* if it represents them with identical probability distributions, i.e. $p(\lambda|P) = p(\lambda|P')$ for all $P \sim P'$.

A similar definition applies to measurements and transformations. Two measurement effects $[k|M], [k|M']$ are operationally equivalent if they give the same results for any preparation, i.e. $[k|M] \sim [k|M']$ if their outcomes can be associated such that for all P one has $p(k|M,P) = p(k|M',P)$. An ontological model is then said to satisfy *measurement noncontextuality* if it equips them with equal response functions, i.e. $p(k|M,\lambda) = p(k|M',\lambda)$ for all $[k|M] \sim [k|M']$. Finally, two transformations T, T' are operationally equivalent, denoted by $T \sim T'$ if $p(k|P,T,M) = p(k|P,T',M)$ for all preparations and measurements P, M. An ontological model is then said to satisfy *measurement noncontextuality* if it equips them with equal transition matrix, i.e. $\mu(\lambda|\lambda',T) = \mu(\lambda|\lambda',T)$ for all $T \sim T'$.

In the paradigm of generalized noncontextuality, it is moreover assumed that probabilistic mixtures of preparations are represented as convex mixtures in the ontological model as well, i.e., probabilistically preparing P_i with probability p_i, leading to measurement statistics $\sum_i p_i p(k|M,P_i)$, is represented as $\sum_i p_i \mu(\lambda|P_i)$, and similarly for

measurements. I name this a *convexity-preserving* assumption, and it is typically taken part of the definition of generalized noncontextuality definition itself.

Definition of generalized noncontextuality. Summarizing, generalized noncontextuality requires that for an operational equivalence of preparations

$$\sum_j \alpha_{P_j} P_j \sim \sum_{j'} \beta_{P_{j'}} P_{j'} \tag{2.16}$$

it holds that

$$\sum_j \alpha_{P_j} \mu_{P_j}(\lambda) = \sum_{j'} \beta_{P_{j'}} \mu_{P_{j'}}(\lambda) \text{ for all } \lambda, \tag{2.17}$$

with $\alpha_{P_j}, \beta_{P_j}$ probability distributions, and for an operational equivalence of measurement effects

$$\sum_{i,m} \alpha_{k|M_i} [k \mid M_i] \sim \sum_{i',k'} \beta_{k'|M_{i'}} [k' \mid M_{i'}] \tag{2.18}$$

it holds that

$$\sum_{i,m} \alpha_{k|M_i} \xi_{k|M_i}(\lambda) = \sum_{i',k'} \beta_{k'|M_{i'}} \xi_{k'|M_{i'}}(\lambda) \text{ for all } \lambda \tag{2.19}$$

with $p_{ki} = \alpha_{k|M_i}, q_{ki} = \beta_{k'|M_{i'}}$ probability distributions over (k, i).

In a finite scenario, the set of correlations in an operational theory satisfying linear constraints derived from generalized noncontextuality forms a convex polytope and characterized by *(generalized) noncontextuality inequalities*. [254, 258]; techniques for addressing nonlinear constraints are yet to be developed [258].

A quantum example violating generalized noncontextuality. The simplest proof that quantum theory manifests *contextuality*, *i.e.*, it does not admit of a generalized noncontextual ontological model, can be constructed in the *simplest nontrivial scenario* [260–262]. This is a prepare-and-measure scenario consisting of two tomographically complete measurements, *e.g.*, the Pauli X and Z measurements on a qubit, and four preparations, *e.g.*, $P_\theta = \cos(\theta/2)\,|0\rangle + \sin(\theta/2)\,|1\rangle$, for $\theta = \pi/4, 3\pi/4, 5\pi/4, 7\pi/4$. These are represented on the Bloch disk in Fig. 2.2. An equal mixture of the preparations $\pi/4, 5\pi/4$ yields the maximally mixed state, and similarly for the two preparations $3\pi/4, 7\pi/4$. Thus, they satisfy the following operational equivalence,

$$\frac{1}{2}p(k|M, P_{\frac{\pi}{4}}) + \frac{1}{2}p(k|M, P_{\frac{5\pi}{4}}) = \frac{1}{2}p(k|M, P_{\frac{3\pi}{4}}) + \frac{1}{2}p(k|M, P_{\frac{7\pi}{4}}) \quad \forall [k|M]. \tag{2.20}$$

However, there does not exist an ontological model where these also provide an ontological identity.

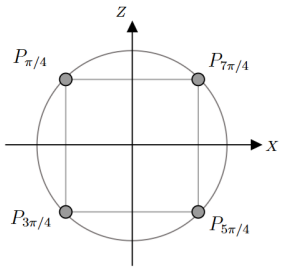

Figure 2.2: The simplest nontrivial scenario [260], involving four preparations (gray dots) and two measurements (the two axes) that provide a maximal violation of the noncontextuality inequalities.

Outcome-deterministic ontological models An ontological model for an operational theory T is called *outcome-deterministic* for measurement M and ontic state λ we have $\xi(k|M,\lambda) \in \{0,1\}$, and outcome-deterministic for a measurement M if it is so for all states λ.

2.3 The measurement problem, Wigner's friend and extensions

In its standard formulation, according to the textbook postulates of Section 2.1.2, quantum theory prescribes two distinct ways of time evolution: (i) the unitary, deterministic evolution of a closed system according to Schrödinger's equation, and (ii) the probabilistic, indeterministic evolution of a system after a measurement upon it, also referred to as the wavefunction collapse. But what causes the wavefunction to collapse? What precisely qualifies as a measurement such that evolution (ii) applies?

This puzzle, also referred to as the measurement problem[5] [3–8] has been a veil of mystery right since the birth of quantum mechanics. A famous illustration of this problem was formulated in 1935 in the form of Schrödinger's cat, where the cat's survival depends on a quantum event, and, according to the unitary Schrödinger evolution, the cat, in a sealed box, evolves into a superposition of dead and alive. However, when we open the box, the cat is found either dead or alive, and the wavefunction has collapsed. But what causes this collapse, and when? Is it the role of the observer, for the cat's wavefunction to collapse only when observed by us, or the chamber, or something else?

[5] Often, one distinguishes several measurement problems; the one described here is also referred to as the 'big' measurement problem [3–5]. Other related problems concern the question of why we do not observe macroscopic superpositions, in whose answer decoherence plays a prominent role [3–6].

Wigner's thought experiment [10], dating back to 1967, is another scenario showcasing the tension of the measurement problem, where an observer referred to as Wigner's friend resides in a sealed lab and measures a quantum system. According to Wigner, a *superobserver*, the friend is ultimately a quantum system too and evolves into a superposition of seeing different outcomes, at odds with the friend's description, and it is only when Wigner opens the friend's lab to ask for their outcome that their descriptions match.

One can argue that this contradiction is only problematic if one believes the quantum state is a real physical state, with only one possible description, irrespective of anything else in the universe [3]. However, the problem then shifts to the question of when, where and how a measurement happens, and which systems would count as measurement devices or observers. Recent results have strengthened the tension in the measurement problem, obtaining a contradiction at the level of outcomes of performed measurements and observers' predictions. More technically, these results consider extensions of Wigner's thought experiment combined with Bell nonlocality. Namely, in a Bell nonlocality scenario not all measurements are performed at once, and several rounds of the experiment are required in order to obtain all relevant joint correlations. In 2015, Brukner did away with this counterfactuality by proposing a thought experiment involving multiple Wigners and their friends, who share a Bell nonlocality state [5], laying the foundations for the Frauchiger–Renner paradox [11] and Local Friendliness no-go theorem [12].

In the rest of this subsection we outline Wigner's friend scenario, after which we present the Frauchiger–Renner paradox and Local Friendliness no-go theorem in Sections 2.4 and 2.5, laying the basis for our work in the next chapters of Part I.

2.3.1 Wigner's friend scenario

The Wigner's friend scenario [10] involves Wigner and his friend, Bob, who resides inside a sealed lab and measures a quantum system S_B. Wigner treats Bob's lab itself as a quantum system, which Wigner, as a superobserver of Bob, may measure and control from the outside. We denote by L_B the system consisting of Bob's lab, including Bob himself but excluding S_B. The setup is represented schematically in Fig. 2.3.

The quantum system S_B is initialised in the state $|\psi\rangle_{S_B} = \alpha\,|0\rangle_{S_B} + \beta\,|1\rangle_{S_B}$ with $\alpha, \beta \neq 0$. The friend measures the system S_B in the computational basis $\{|0\rangle_{S_B}, |1\rangle_{S_B}\}$, labelling the outcomes $0, 1$. As a *superobserver* for his friend, Wigner models Bob's

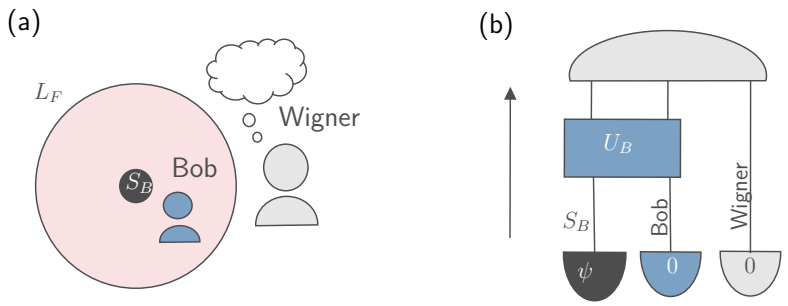

Figure 2.3: (a): Schematic sketch of Wigner's friend scenario. Bob performs a measurement on a system S_B in a sealed lab L_B. Wigner is a superobserver of his friend Bob, modelling Bob's measurement unitarily as U_B. The friend sees a definite single-valued outcome, but Wigner describes his friend as being in a superposition of having obtained different outcomes. (b): Circuit representation of a Wigner's friend scenario.

measurement of the system S_B as a unitary evolution of the joint quantum system $S_B L_B$[6]:

$$|\psi\rangle_{S_B} |r\rangle_{L_B} \xmapsto{U_B} \alpha |00\rangle_{S_B L_B} + \beta |11\rangle_{S_B L_B} , \tag{2.21}$$

where $|0\rangle_{L_B}$ and $|1\rangle_{L_B}$ correspond to the two possible outcomes as recorded in Bob's memory and elsewhere in his lab, and $|r\rangle_{L_B}$ denotes the initial ('ready') state of his lab. From Wigner's perspective, the system $S_B L_B$ is in a superposition of 'S_B is in state $|0\rangle$ and Bob has recorded the measured outcome 0' and 'S_B is in state $|1\rangle$ and Bob has recorded the measured outcome 1'. Note that the unitary evolution U_B is described (on the relevant degrees of freedom) as a CNOT gate. The friend sees a definite single-valued outcome, while Wigner would describe his friend as being in a superposition of states corresponding to different outcomes.

Both claims can in principle be tested, as in Deutsch's version of the Wigner's friend argument [263]; see also Ref. [264, Section 4]. Indeed, the fact that the friend has obtained a definite, single-valued outcome can in principle be known to Wigner, namely by the friend passing a note to Wigner stating 'I see a definite single-valued outcome' (but crucially without giving away any information about what the outcome actually is). Wigner's description of the friend as being in a superposition state can also in principle be tested. Namely, as a superobserver, Wigner can perform state tomography on the entire friend's lab. One way for Wigner to perform such state tomography is the following [264]. As a superobserver, Wigner can *undo* the friend's measurement, by applying the inverse unitary $U_B^\dagger$ to the friend's lab and the system, thereby resetting them back to their initial

[6]We will often omit tensor product when it is clearly implied, i.e. here $S_B L_B$ is an abbreviation of $S_B \otimes L_B$.

state. Thus, by applying $U_B^\dagger$ and verifying whether the obtained initial state of $S_B L_B$ is indeed $|\psi\rangle_{S_B}|0\rangle_{L_B}$ Wigner can verify his superposition assignment. Note that Wigner's *undoing* of the friend's measurement is only effective, i.e. only resets the lab back to its initial state, if no information about the friend's measurement outcome escapes the lab.

An actual human playing the role of the friend may be practically infeasible, but an artificially intelligent program on a future quantum computer may be a reasonable physical realization of such an observer [265].

This thought experiment shows the incompatibility between the existence of superobservers and an absolute, observer-independent notion of collapse [12, 30, 263]. However, even if one takes the wave function collapse not to be absolute, the possibility remains that the underlying fact of the friend's observed outcome be absolute [12]. To see this, suppose Wigner and his friend run the protocol many times. In some rounds Wigner verifies his superposition assignment of Eq. (2.21) for the state of his friend Bob's lab, while in other rounds Wigner directly asks his friend Bob for the outcome. In that case, the two may disagree on (the timing of) the occurrence of collapse, but they agree on the underlying fact, namely which outcome the friend observed. While not asking his friend, Wigner may describe his friend as being in a superposition, but this does not rule out the possibility of there being an absolute fact regarding which outcome the friend obtained, even if unknown to or hidden from Wigner.

This is where extended Wigner's friend scenarios, involving more than one superobserver, come into play. Since Brukner's scenario [5], many different extended Wigner's friend scenarios have been proposed [11–16, 18, 20, 23, 25–32, 266]. Among them, the Frauchiger–Renner paradox [11] and Local Friendliness no-go theorem [12, 13] have gained much attention.

Terminology: superobservers and classical agents. For reasons of clarity, we briefly define what we mean by an agent, a measurement and introduce terminology regarding (super)observers and (classical) agents.

Definition of an agent and measurement. An agent is any system that can obtain information through measurements, and store information. In quantum information theory using the Hilbert space formalism, the ability to store information means that an agent Bob has a personal preferred orthogonal basis in which he can store information. Storing information in these orthogonal 'knowledge' states makes sure his information states are perfectly distinguishable for him. The basis must be a preferred one because

not every quantum basis can be seen as encoding information for a single observer [267, 268]. By a measurement an interaction of the agent Bob with a system S is meant, that increases the information in his memory (i.e. in his knowledge states, orthogonal states in some (personal) preferred basis). The agent being able to obtain information thus means that the agent must be able to perform some measurements, and store the outcome in his knowledge states[7].

Terminology regarding superobservers and classical agents. We will use the term observer and agent interchangeably, and by a superobserver, such as Wigner, we mean an observer who has extreme quantum control over another observer (their 'friend'). As a result, a superobserver models the evolution of their friend within their sealed lab quantumly, i.e. as a unitary, including measurements performed by this friend. Within a given measurement protocol, by a classical agent we mean an agent who need not be modeled quantumly by anyone else in the protocol.

2.4 The Frauchiger–Renner paradox

In this subsection we review the Frauchiger–Renner (FR) paradox. We observe that the FR paradox (i) can be strengthened by a slight modification in the reasoning stage, and (ii) is based on the nonlocal Hardy model. These observations suggest a stronger FR-like paradox that we describe in Chapter 3, where we also carefully discuss the underlying assumptions.

2.4.1 Entanglement version of the FR paradox

We sketch the entanglement version of the FR paradox as outlined in Ref. [24], credited to Lluis Masanes. For the original prepare-and-measure version of the FR paradox we refer to Refs. [11, 24].

The scenario is depicted in Fig. 2.4(a). It involves two superobservers, Ursula and Wigner, each with their own 'friend', Alice and Bob respectively. Each of the friends reside in a sealed lab with access to quantum systems S_A and S_B respectively, which are

[7] Storing an outcome in a preferred basis might lead to production of entropy as in practice it corresponds to a state preparation, see for example [269], and thus might require suitable entropic conditions and an (agential) thermodynamic arrow of time. A preferred basis might also arise from interactions.

initialised in a shared entangled state. As before, L_A denotes Alice's lab including Alice herself but excluding the system S_A, and similarly L_B for Bob. All the agents – Alice, Bob, Ursula, and Wigner – perform measurements, obtaining outcomes $a, b, u, w \in \{0, 1\}$ respectively.

Measurement protocol. The measurement protocol, sketched in circuit form in Fig. 2.4(b), proceeds as follows;

1. **Initialisation.** The system $S_A S_B$, shared between Alice and Bob, starts in the specific entangled state $|\psi_0\rangle_{S_A S_B} = \frac{1}{\sqrt{3}} (|00\rangle + |10\rangle + |11\rangle)_{S_A S_B}$. The initial states of Alice's and Bob's labs are labelled by $|r\rangle_{L_A}$ and $|r\rangle_{L_B}$, respectively. The initial state of $S_A L_A S_B L_B$ is given by[8]

$$\left|\psi^{t=1}\right\rangle_{S_A L_A S_B L_B} = \frac{1}{\sqrt{3}} (|0r0r\rangle + |1r0r\rangle + |1r1r\rangle)_{S_A L_A S_B L_B} . \tag{2.22}$$

2. **Alice and Bob measure.** Alice measures her system S_A in the $\{|0\rangle, |1\rangle\}$ basis, and Bob similarly measures his system S_B. The unitary description of these measurements, as in Eq. (2.21), leads to the overall state

$$\begin{aligned} \left|\psi^{t=2}\right\rangle_{S_A L_A S_B L_B} &= (U_A \otimes U_B) \left|\psi^{t=1}\right\rangle_{S_A L_A S_B L_B} \\ &= \frac{1}{\sqrt{3}} (|0000\rangle + |1100\rangle + |1111\rangle)_{S_A L_A S_B L_B} . \end{aligned} \tag{2.23}$$

   where U_A and U_B are the unitaries describing Alice's and Bob's measurements, respectively.

3. **Ursula and Wigner supermeasure.** Finally, the superobservers Ursula and Wigner respectively measure the systems $S_A L_A$ and $S_B L_B$ in a basis with vectors[9]

$$|\text{ok}\rangle = \frac{1}{\sqrt{2}} (|00\rangle - |11\rangle), \qquad |\text{fail}\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle). \tag{2.24}$$

   Note that these measurements can be seen as the composition of (i) undoing the friend's measurements by applying the inverse unitaries $U_A^\dagger, U_B^\dagger$, followed by (ii) an X-measurement on S_A, S_B, respectively.

[8] For notational ease, we will often omit the tensor product symbol: $|00\rangle_{AB} = |0\rangle_A |0\rangle_B = |0\rangle \otimes |0\rangle$.

[9] Note that a basis requires two more orthonormal vectors, but their corresponding outcomes have probability zero, so we do not need to specify the remaining vectors.

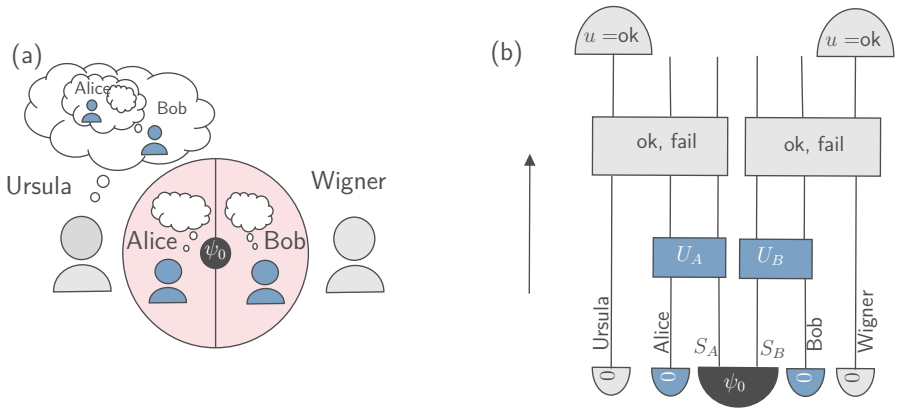

Figure 2.4: (a): Sketch of the FR paradox, where Ursula reasons about Bob who reasons about Alice who reasons about Wigner. (b): Circuit of the entanglement version of the FR paradox, as outlined in Ref. [24], with the initial state of $S_A \otimes S_B$ given by $|\psi_0\rangle_{S_A S_B} = \frac{1}{\sqrt{3}}(|00\rangle + |10\rangle + |11\rangle)_{S_A S_B}$.

The protocol is post-selected on both Ursula and Wigner obtaining the outcome ok, denoted $u = \text{ok}$ and $w = \text{ok}$. Observe that this happens with nonzero probability:

$$|\langle \text{ok}|_{S_A L_A} \otimes \langle \text{ok}|_{S_B L_B} \left|\psi^{t=2}\right\rangle_{S_A L_A S_B L_B}|^2 \tag{2.25}$$

$$= |\frac{1}{2\sqrt{3}}(\langle 0000| - \langle 0011| - \langle 1100| + \langle 1111|)(|0000\rangle + |1100\rangle + |1111\rangle)|^2 \tag{2.26}$$

$$= |\frac{1}{2\sqrt{3}}|^2 = \frac{1}{12} > 0. \tag{2.27}$$

Reasoning steps. After a successful post-selected run of the protocol, Ursula reasons (i) about Bob's outcome b, (ii) about Bob's conclusion for Alice's outcome a, and (iii) about Bob's reasoning about Alice's reasoning about Wigner's outcome w. Assuming that each agent uses the Born rule to make predictions, Ursula, using $u = \text{ok}$, reasons that Bob must have obtained $b = 1$, and thus that he would have predicted that $a = 1$, so that Alice would have concluded that $w = \text{fail}$. In summary,

$$u = \text{ok} \;\Rightarrow\; b = 1 \;\Rightarrow\; a = 1 \;\Rightarrow\; w = \text{fail},$$

contradicting the initial post-selection on $u = \text{ok}, w = \text{ok}$. We give additional details in the next section where we show how the argument may in fact be strengthened.

2.4.2 A slightly stronger modified FR paradox

We provide a stronger variant of the FR paradox in which only *classical*[10] agents reason. This technique will be used in the GHZ–FR paradox of Chapter 3 as well.

The measurement protocol is identical to that of the FR paradox. But an additional agent, Zeno,[11] who does not perform any measurement, makes a prediction for Alice's and Bob's outcomes. Ursula and Wigner, after obtaining their outcomes, reason about the outcomes obtained by Alice and Bob, respectively. Ursula, and Wigner hand their predictions to Zeno, who derives a contradiction.[12]

Post-selecting on $u = \text{ok}, w = \text{ok}$, which has a non-zero probability of occurring, we have the following predictions from applications of the Born rule:

- Ursula predicts that $b = 1$ by excluding $b = 0$ since

$$\langle \text{ok}|_{S_A L_A} \langle 00|_{S_B L_B} \left|\psi^{t=2}\right\rangle_{S_A L_A S_B L_B} = 0\,.$$

- Wigner predicts that $a = 0$ by excluding $a = 1$ since

$$\langle 11|_{S_A L_A} \langle \text{ok}|_{S_B L_B} \left|\psi^{t=2}\right\rangle_{S_A L_A S_B L_B} = 0$$

- Zeno predicts that $a = 0, b = 1$ can never occur since

$$\langle 00|_{S_A L_A} \langle 11|_{S_B L_B} \left|\psi^{t=2}\right\rangle_{S_A L_A S_B L_B} = 0\,.$$

Ursula and Wigner classically communicate their predictions to Zeno. Upon combining the predictions made by Ursula and Wigner with his own, Zeno finds a contradiction. Thus, all reasoning and predicting is done by classical agents. Furthermore, each application of the Born rule refers to measurement outcomes that are obtained concurrently: its predictions could in principle be tested by interrupting the protocol before one or both supermeasurements.

2.4.3 Logical non-locality powers the FR paradox

In Section 2.4.2 we have seen how actually a contradiction can be obtained by having only *classical* agents reason in the FR protocol. The second crucial observation about the

[10] For a definition of a *classical* agent, see Section 2.3.1.

[11] The additional agent is needed here to reason about the joint outcomes of the friends Alice and Bob, but will not be necessary in the case of the GHZ–FR paradox of Section 3.1.

[12] We note that Ref. [270] gives a presentation of the reasoning steps of the FR paradox that comes close to the one here.

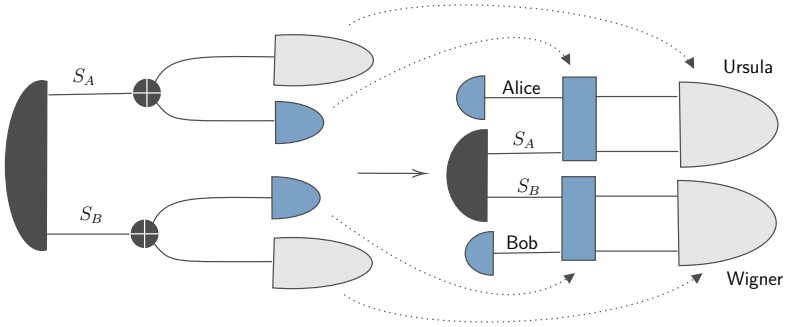

Figure 2.5: Mapping a Bell non-locality scenario (left) to an FR-like extended Wigner's friend scenario (right). In the Bell scenario each party can choose (pictured by a crossed circle) one out of two (incompatible) measurements, for example in the Z basis $\{|0\rangle, |1\rangle\}$ or in the X basis $\{|+\rangle, |-\rangle\}$. In the corresponding extended Wigner's friend scenario, for each Bell party, one of the measurements, for example the computational basis $|0\rangle, |1\rangle$ measurement, is performed by an observer and modelled unitarily as U, whereas the other measurement is performed by the corresponding superobserver in an entangled basis $U|\pm\rangle_{S_F}|0\rangle_{L_F}$.

FR paradox is to recognise that it is underpinned by the Hardy model [271], a proof of "nonlocality without inequalities". This approach has a logical flavour in that nonlocality is witnessed in terms of *possibilities* alone rather than requiring full knowledge of the *probabilities*. In fact, Frauchiger and Renner [11] explicitly stated that their protocol is motivated by Hardy's paradox, a connection that has also been emphasized in subsequent works such as Refs [11, 23, 24, 31, 272–274].

The Hardy model is a specific quantum realisation of the simple $(2, 2, 2)$ Bell scenario. This nonlocality scenario involves two parties, each able to choose between two measurement settings, with two outcomes 0 and 1. The Hardy model, on which the FR paradox is based, arises from a specific choice of initial shared state and local measurements in the $(2, 2, 2)$ Bell scenario. The two parties share a system in the entangled state $1/\sqrt{3}(|00\rangle + |10\rangle + |11\rangle$, which each can choose to measure locally in either the $\{|0\rangle, |1\rangle\}$ or the $\{|+\rangle, |-\rangle\}$ basis, known respectively as the Z- and the X-basis.[13]

Fig. 2.5 depicts a mapping between the $(2, 2, 2)$ Bell scenario and the extended Wigner's Friend scenario of the FR paradox.

The key idea is better explained by focusing on a *single* party. That is an experimenter who, in each run of the experiment, chooses between two (incompatible) dichotomic measurements to perform on their part of a shared system. Such a situation can be translated into a Wigner's friend scenario: the friend first performs one of the measurements, and then the superobserver Wigner undoes this measurement and performs

the other. The choice of measurement setting is thus translated to a choice between adopting the point of view of Wigner or that of the friend. Applying the same idea to each party, the translation extends to the $(n, 2, 2)$ Bell scenario for any number n of parties.

As discussed in Section 2.2, Bell nonlocality scenarios are a particular case of contextuality scenarios, where maximal measurement contexts correspond to a choice of one measurement for each party. In the $(2, 2, 2)$ Bell scenario there are four dichotomic measurements, two for each party. We choose to label Alice's measurements as A and U and Bob's as B and W, foreshadowing the link to the FR scenario. The maximal measurement contexts are then

$$\{A, B\}, \quad \{U, B\}, \quad \{A, W\}, \quad \{U, W\}.$$

Using the Born rule we calculate the (two-party) joint outcome probabilities, yielding a probability distribution over joint outcomes for each context. In fact, in this case, the *possibilistic* version of the Born rule suffices, which only distinguishes between *possible* (probability greater than zero) and *impossible* (probability zero) events. Table 2.1 indicates such possibilistic empirical data for the specific choice of state and measurements in the Hardy model: each row is labelled by a measurement context and indicates the support of the probability distribution on the joint outcomes for that context; see e.g. [42, 275].

Contextuality, of which non-locality is a special instance, occurs when there is no global probability distribution (on assignments of outcomes to all the measurements in the scenario) that marginalises to the empirical probability distributions within each measurement context. The Hardy model [271] witnesses *logical* non-locality such that non-locality can be ascertained from the possibilistic information collected in Table 2.1 alone. The argument corresponds precisely to the logical reasoning chain in the FR paradox. Specifically, the joint outcome assignment $\{u \mapsto 0, w \mapsto 0\}$, which occurs with positive probability, cannot be extended to a global assignment for all a, b, u, w which does not restrict to an impossible event in some other context (see caption of Table 2.1 for details). We note that the troublesome global assignments in the contextuality setting correspond to statements about (absolute) outcomes in FR-like scenarios.

[13] We regard $|0\rangle$, $|+\rangle$ as corresponding to the outcome label 0 and $|1\rangle$, $|-\rangle$ to the label 1. Note that these are, respectively, the $+1$ eigenvectors and the -1 eigenvectors of the Z and X matrices. Often, when regarding the Z or X matrices as measurement operators, one considers outcomes valued ± 1. Here, we use the bijection $+1 \mapsto 0$, $-1 \mapsto 1$. Particularly relevant for the GHZ–Mermin argument, this bijection actually yields a group isomorphism between $\{+1, -1\}$ under multiplication and $\{0, 1\}$ under addition modulo 2, an instance of the isomorphism between the multiplicative group of complex n-th roots of unity and the additive group of integers modulo n.

	$0,0$	$0,1$	$1,0$	$1,1$
A,B	1	0	1	1
A,W	1	1	0	1
U,B	0	1	1	1
U,W	1	1	1	1

Table 2.1: Support of joint outcome probability distributions for each context in the Hardy model or the FR paradox. The rows are labelled by the four measurement contexts $\{A,B\},\{A,W\},\{U,B\},\{U,W\}$ and columns by the joint outcomes $(0,0),(0,1),(1,0),(1,1)$, the combination of which gives an event: a specific joint outcome for a specific context of measurements. The entries in each cell of this possibility table have two possible values: 1 denotes an event being *possible* while 0 denotes an event being *impossible*, i.e. having zero probability. Note that in the FR case the outcomes u and w are labelled by ok, fail instead of $0,1$. This table is logically contextual: there is no consistent global assignment of values to A,B,U,W that extends the possible event $\{U\mapsto 0, W\mapsto 0\}$ (in the bottom left); starting from this partial assignment, row 3 forces $B\mapsto 1$, which used in row 1 gives $A\mapsto 1$, which used in turn in row 2 yields $W\mapsto 1$, contradicting the initial assignment.

The mapping from a Bell scenario to an extended Wigner's friend scenario as in Fig. 2.5 provides a recipe for producing extended Wigner's friend paradoxes based on nonlocal models. We will exploit this recipe in the next section to produce an FR-like paradox based on the GHZ–Mermin model [276, 277].

2.5 The Local Friendliness no-go theorem

The Local Friendliness scenario is similar to the Frauchiger–Renner protocol, with the difference that the two superobservers in the Local Friendliness scenario are each given a choice x,y whether to ask their friend for their measurement outcome or to perform their own measurement which involves undoing the measurements of the friends. In this way, the measurement statistics $\wp(a,b|x,y)$ are obtained by Alice and Bob.[14] Instead of a paradox from agents reasoning about each other as in the FR paradox, a contradiction is now obtained with the Local Friendliness assumptions of Absoluteness of Observed Events and Local Agency.

[14] Similarly to Refs. [12, 21, 22], when working in the context of the Local Friendliness no-go theorem or related results, we introduce a notational difference between empirical probability distributions that can be obtained by a single agent from the collected data, denoted by $\wp$, and theoretical probability distributions required to exist given certain assumptions, but potentially not observable by any agent, denoted by p.

An unfortunate convention in literature is that regarding the names of the involved agents different conventions are used in the literature on the Local Friendliness no-go theorem compared to the Frauchiger–Renner paradox. Specifically, the superobservers of the FR paradox are now called Alice and Bob in the LF scenario, as opposed to Ursula and Wigner, and the friends are called Charlie and Debbie. Whenever we will be concerned with the Local Friendliness results, we will thus use these conventions to be in line with literature.

We now outline the LF scenario and no-go theorem in some more detail, following Ref. [22] closely.

2.5.1 Local Friendliness scenarios

The experimental setup, illustrated in Fig. 2.6, is described as follows. We will present a theory-independent way of specifying the scenario, which we refer to as the (N_A, N_B)-setting LF scenario, while also outlining how the scenario for a quantum violation goes (taking $N_A = 2 = N_B$ for this).

Charlie and Debbie, spacelike separated, share a bipartite system in the state ρ_{RS}. They perform measurements denoted with C, D on their respective subsystems, R, S, obtaining outcomes $c, d \in \{0, 1\}$. In quantum theory, the measurements are modelled by unitaries U_C, U_D, respectively. After Charlie has performed his measurement, Alice makes a choice $x \in \{1, \ldots, N_A\}$ with $N_A \geq 2$. If $x = 1$ she asks Charlie for his outcome and sets her outcome equal to his, revealing his outcome to the world. If $x \neq 1$ she applies another operation and measurement to Charlie's lab, choosing out of $N_A - 1$ measurements labelled by $x \in \{2, \ldots, N_A\}$. In particular, for the quantum scenario that will provide a violation of the LF assumptions, we can take $N_A = 2$ and for $x = 2$ Alice acts as a superobserver and undoes Charlie's measurement by applying $U_C^\dagger$, thereby erasing the outcome record of c; subsequently she performs a measurement denoted with A on R, obtaining outcome $a \in \{0, 1\}$. Similarly, Bob chooses $y \in \{1, \ldots, N_B\}$. For $y = 1$, he takes Debbie's outcome to be his outcome. For $y \neq 1$ he applies another operation and measurement to Debbie's lab, choosing out of $N_B - 1$ measurements. For the quantum scenario violating LF, we can take $N_B = 2$ and for $y = 2$ Bob undoes Debbie's measurement and subsequently performs a measurement denoted with B on S, obtaining outcome $b \in \{0, 1\}$.

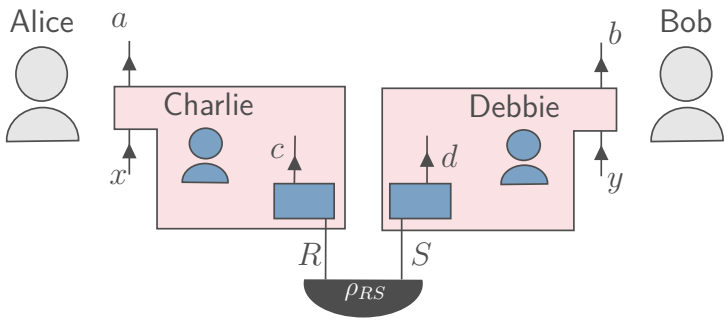

Figure 2.6: Local Friendliness setup [27].

2.5.2 No-go theorem

Given the setup, we state the required assumptions of Absoluteness of Observed Events and Local Agency, whose conjunction is also referred to as 'Local Friendliness' (LF).[15]

Assumption 1 (Absoluteness of Observed Events (AOE))**.** An observed event is an absolute single event, not relative to anything or anyone.

This assumption can be motivated as an instance of realism and, in the LF scenario, it implies that there exist distributions $p(a, b, c, d|x, y)$ from which one can obtain the empirical probabilities, $\wp(a, b|x = 2, y = 2), \wp(a, d|x = 2, y = 1), \wp(c, b|x = 1, y = 2), \wp(c, d|x = 1, y = 1)$ via marginalisation.[16]

Assumption 2 (Local Agency)**.** The only relevant events correlated with an intervention are in its future light cone.

In the LF scenario, Local Agency implies, for example, that $p(c, b|x, y = 2) = p(c, b|y = 2)$, as none of c, b are obtained in the future of the choice x.

We can now state the LF no-go theorem [12–14].

Theorem 1 (LF no-go theorem)**.** *If a superobserver can perform arbitrary quantum operations on an observer and its environment, then no physical theory can satisfy Local Friendliness.*

[15] Regarding terminology, the no-go theorem against the Local Friendliness assumptions as in Theorem 1 is typically referred to as the Local Friendliness no-go theorem, and scenarios used to prove it as Local Friendliness scenarios.

[16] To argue that Alice's result of asking Charlie for his outcome ís Charlie's outcome, an additional "tracking assumption" [27] is introduced, i.e. $p(a \mid c, x = 1) = \delta_{a,c}$, in that 'Alice's outcome tracks Charlie's exactly in the relevant runs of the protocol', and similarly for Bob and Debbie. For a discussion on why this is considered a minimal assumption, see [27].

Proof. Let us consider the LF scenario of figure 2.6 with the following specifications. Charlie and Debbie share the bipartite entangled state $(|01\rangle - |10\rangle)/\sqrt{2}$, and perform an $(X+Z)/\sqrt{2}$ and Z measurement, respectively. When $x=2$, Alice undoes Charlie's measurement and performs an $(Z-X)/\sqrt{2}$ measurement; when $y=2$, Bob undoes Debbie's measurement and performs an X measurement (any set of four measurements and an entangled state that produce correlations violating a Bell inequality would also work [17, 21]). There are four empirical distributions:

$$\begin{aligned} &\wp(c,d \mid x=1,y=1),\\ &\wp(c,b \mid x=1,y=2),\\ &\wp(a,d \mid x=2,y=1),\\ &\wp(a,b \mid x=2,y=2). \end{aligned} \tag{2.28}$$

By AOE and Local Agency, we can rewrite these empirical correlations as follows,

$$\wp(c,d \mid x=1,y=1) = p(c,d|x=2,y=2), \tag{2.29}$$

$$\wp(c,b \mid x=1,y=2) = p(c,b|x=2,y=2), \tag{2.30}$$

$$\wp(a,d \mid x=2,y=1) = p(a,d|x=2,y=2), \tag{2.31}$$

$$\wp(a,b \mid x=2,y=2) = p(a,b|x=2,y=2). \tag{2.32}$$

AOE is used to consider joint probability distributions when c or d are involved and possibly erased. To obtain Eq. (2.30) we first appeal to an instance of Local Agency that enforces no-superdeterminism, $p(c|xy)=p(c)$, yielding $p(c|x=1,y=2)=p(c|x=2,y=2)$; we can then apply the rule of conditional probability and consider another instance of Local Agency expressing that neither b nor c occur in the future light cone of x: $p(b|c,x,y){=}p(b|c,y)$, thus obtaining

$$\begin{aligned} &\wp(c,b|x{=}1,y{=}2) = p(c|x{=}1,y{=}2)p(b|c,x{=}1,y{=}2)\\ &= p(c|x{=}2,y{=}2)p(b|c,x{=}2,y{=}2) = p(c,b|x{=}2,y{=}2). \end{aligned}$$

Similar instances of Local Agency lead to Eq. (2.31). Eq. (2.29) derives from an instance of Local Agency enforcing no-superdeterminism, $p(c,d|x,y)=p(c,d)$.

By assuming AOE, we can assert that all the correlations in Eqs. (2.29), (2.30) ,(2.31) and (2.32) can be obtained from a single joint probability distribution $p(a,b,c,d|x=2,y=2)$ via marginalization. However, given the choices of preparations and measurements made, by Fine's theorem [42, 247], such distribution cannot exist. Indeed, these give exactly the empirical correlations that lead to a maximal violation of the Bell inequalities. ■

In terms of rigor, the LF no-go theorem has the same status as Bell's theorem, since it is precise, theory-independent, and – at least in principle – empirically testable [265].

— 3 —

A refined Frauchiger–Renner paradox based on strong contextuality

in this chapter we present a refined version of the Frauchiger–Renner paradox, where the consistency assumption is weakened significantly. Namely, a contradiction is derived when agents who meet and communicate declare their Born rule predictions to each other. In contrast, in the FR paradox, an agent needs to trust the reasoning and predictions of an agent in a sealed lab with whom they cannot meet and communicate about their predictions. This is particularly interesting as already some classical paradoxes such as the Sleeping Beauty problem violate FR's consistency condition, but do not violate our weakened consistency requirement. We discuss the Sleeping Beauty problem and a particularly simple resolution as an intermezzo at the end of this Part, in Chapter 6. Our result implies that the route taken to tackle the FR paradox by rejecting its consistency requirement needs a more radical resolution, as it needs to reject even our weakened consistency condition. We will shortly discuss potential ways out of the (refined) FR paradox, and argue how, when taking the route of rejecting the Absoluteness of Observed Events, our paradox suggests a relational principle which we name the 'Relativity of Observed Events'.

A key objective of this chapter is to rigorously define the underlying principles and assumptions. To aid clarity, assumptions and their references throughout this chapter are highlighted using color and hyperlinks. This chapter follows Ref. [19], nearly verbatim.

3.1 Refining the Frauchiger–Renner paradox

In this section we present the protocol of our refined version of the Frauchiger–Renner (FR) paradox, based on the strongly nonlocal GHZ–Mermin model. In the previous chapter we have seen how the Bell nonlocality scenario with two parties with each two measurements underlies the FR paradox (and the Local Friendliness no-go theorem). We present an FR-like paradox but based on a model with three parties that has a stronger form of nonlocality. As a consequence, a contradiction is derived in each round of the protocol, doing away with the post-selection step of the FR paradox. Moreover, as already hinted at in Section 2.4.2, the reasoning steps are formulated differently from the original FR scenario, leading to a weaker consistency assumption to derive a stronger no-go theorem.

3.1.1 Strong contextuality: the GHZ–Mermin argument

Strong contextuality. As recapped in Section 2.2.1, the sheaf-theoretic approach to non-locality and Kochen–Specker (KS) contextuality [42, 275, 278] introduced a hierarchy of contextuality of increasing strength: probabilistic, logical (or possibilistic), and strong contextuality. In quantum theory, each of these levels of the hierarchy is exemplified by a well-known model:

$$\text{Bell–CHSH (probabilistic)} \prec \text{Hardy (logical)} \prec \text{GHZ–Mermin (strong)}.$$

The contextual model underlying the FR paradox is the logically contextual Hardy model: contextuality is witnessed at the level of the *support* of probability distributions. Namely, there exists a *possible* assignment of outcomes to a context which cannot be extended to a global assignment (to all measurements) consistent with the model, i.e. whose restriction to each context is deemed possible.

Nevertheless, in the Hardy model, there exists a local assignment that can be extended globally; e.g. the global assignment $\{a \mapsto 1, b \mapsto 1, u \mapsto 1, w \mapsto 1\}$ is consistent with the observed possibilities in Table 2.1. *Strong contextuality* holds when not a single such global assignment exists.[1]

[1]This property coincides with *maximal contextuality* in the sense that the fraction of the model that can be explained by noncontextual hidden variables is zero, or that the algebraic (no-signalling) maximum of a noncontextuality inequality violation is attained [279].

The GHZ–Mermin model. A paradigmatic example of quantum-realizable strong contextuality is given by the GHZ–Mermin model [276, 280], a specific realisation of the Bell scenario with three parties, each having two measurements with each two measurement outcomes. Following the presentation in Ref. [275], we consider three parties, A, B, and C, that share an entangled three-qubit system prepared in the GHZ state,

$$\left|\psi^{\mathsf{GHZ}}\right\rangle_{ABC} = \frac{1}{\sqrt{2}}\left(|000\rangle + |111\rangle\right)_{ABC}.$$

Each party performs either an X or a Y measurement on their qubit, i.e. a measurement in either the $\{|+\rangle, |-\rangle\}$ or the $\{|+i\rangle, |-i\rangle\}$ basis, where $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$ and $|\pm i\rangle = (|0\rangle \pm i\,|1\rangle)/\sqrt{2}$. This leads to measurements $X_A, Y_A, X_B, Y_B, X_C, Y_C$ with outcomes $x_A, y_A, x_B, y_B, x_C, y_C \in \{0, 1\}$. The outcomes corresponding to $|+\rangle$ and $|+i\rangle$ are labeled 0, those corresponding to $|-\rangle$ and $|-i\rangle$ are labeled 1.

A measurement context in this scenario consists of a choice of measurement for each party. We abbreviate the context $\{X_A, X_B, X_C\}$ as XXX and analogously for the others. Using the Born rule to distinguish possible from impossible events for measurement contexts XXX, XYY, YXY and YYX, one obtains Table 3.1. Each row is labeled by a measurement context and indicates the support of a probability distribution on the respective joint outcomes; see e.g. [42, 275]. The outcomes satisfy the parity equations

$$x_A \oplus x_B \oplus x_C = 0, \quad x_A \oplus y_B \oplus y_C = 1, \quad y_A \oplus x_B \oplus y_C = 1, \quad y_A \oplus y_B \oplus x_C = 1, \tag{3.1}$$

where $\oplus$ denotes addition modulo 2.[2] Assigning definite outcome values to all six measurements simultaneously, summing these four equations yields $0 = 1$, a contradiction. Therefore, no global assignment of outcomes to all six measurements is compatible with the possible observations for all measurement contexts – the GHZ–Mermin model is strongly contextual, and as it is a nonlocality scenario, strongly nonlocal.

3.1.2 Description of the GHZ–FR paradox

Mapping Bell nonlocality scenarios to extended Wigner's friend scenarios, as described in Section 2.4.3, we build an FR-like paradox based on the GHZ–Mermin model. The strong contextuality of this nonlocal model does away with the need for post-selection in the GHZ–FR paradox.

The scenario is depicted in Fig. 3.1(a). It involves three observer-superobserver pairs – Alice and Ursula, Bob and Valentina, and Charlie and Wigner – corresponding to the

[2] This argument is also often presented with multiplication of ± 1-labeled outcomes; see Footnote 13.

	$0,0,0$	$0,0,1$	$0,1,0$	$0,1,1$	$1,0,0$	$1,0,1$	$1,1,0$	$1,1,1$
XXX	1	0	0	1	0	1	1	0
XYY	0	1	1	0	1	0	0	1
YXY	0	1	1	0	1	0	0	1
YYX	0	1	1	0	1	0	0	1

Table 3.1: Table with outcome (im)possibilities for measurements by three parties who share a GHZ state. The rows correspond to the support of probability distributions over the outcomes (columns 1–8) that arise when the measurements specified in the first column are performed (XXX, XYY, YXY, YYX).

three parties in the GHZ–Mermin model. The friends each reside in a sealed lab with access to quantum systems S_A, S_B, and S_C respectively, which are initialized in a shared entangled state. As before, L_A denotes Alice's lab including Alice herself but excluding the system S_A, and similarly L_B and L_C for Bob and Charlie.

Alice, Bob, Charlie and their respective superobservers Ursula, Valentina, and Wigner perform measurements with outcomes $a, b, c, u, v, w \in \{0, 1\}$. A circuit representation of the measurement protocol is shown in Fig. 3.1(b). First, the observers A, B, C perform Y-measurements on their part of a GHZ state. Then, Ursula performs an X-like measurement on Alice and her qubit, and writes down her predictions for the outcomes of B, C. One intuition for why it might be natural for Ursula to formulate predictions about b, c is in case she does not know about Valentina's and Wigner's actions (or existence even); in that case she believes to still be able to gather Bob's outcome record, for example by asking him for his outcome. Valentina and Wigner act analogously. Once the protocol has been carried out, Ursula, Valentina and Wigner meet up and share their predictions, deriving a contradiction from them, as we will explain below.

We first describe the measurement protocol in more detail, after which we present the reasoning steps through which they derive a contradiction.

Measurement protocol. The measurement protocol, sketched in circuit form in Fig. 3.1, proceeds as follows:

1. **Initialisation.** The system $S_A S_B S_C$, shared between Alice, Bob, and Charlie, starts in the GHZ state $|\psi^{\mathsf{GHZ}}\rangle_{S_A S_B S_C} = \frac{1}{\sqrt{2}} (|000\rangle + |111\rangle)_{S_A S_B S_C}$. The initial states of the three labs are labeled by $|r\rangle_{L_A}$, $|r\rangle_{L_B}$, $|r\rangle_{L_C}$ (with r from 'ready'). The initial state of $S_A L_A S_B L_B S_C L_C$ is thus given by

$$\left|\psi^{t=1}\right\rangle_{S_A L_A S_B L_B S_C L_C} = \frac{1}{\sqrt{2}} (|0r0r0r\rangle + |1r1r1r\rangle)_{S_A L_A S_B L_B S_C L_C} .$$

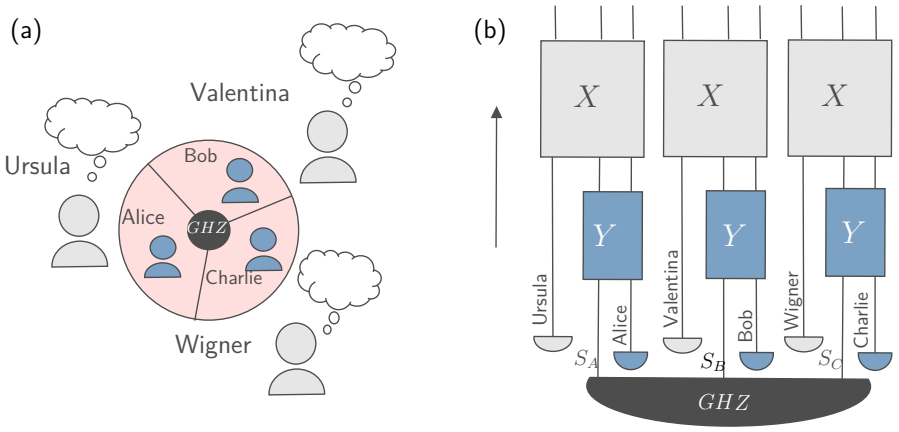

Figure 3.1: (a): Pictorial representation of the scenario GHZ–FR protocol, with superobservers U, V, W and respective observers A, B, C, each having access to part of a GHZ state. (b): The GHZ–FR protocol in circuit form. Alice, Bob and Charlie each have access to one part of a three-qubit GHZ state and measure their qubit in the Y-basis. Ursula, Valentina and Wigner (red), superobservers for Alice, Bob and Charlie, respectively, each perform an X-like measurement on Alice, Bob and Charlie and their qubits.

2. **Alice, Bob, and Charlie measure.** Alice, Bob and Charlie perform a measurement in the Y-basis $|\pm i\rangle$ on their respective qubit S_A, S_B, S_C, with outcomes $0, 1$. Superobserver Ursula describes Alice's measurement by the unitary evolution U_A that, similar to Eq. (2.21) must satisfy:

$$|+i\rangle_{S_A} |r\rangle_{L_A} \overset{U_A}{\longmapsto} |+i\rangle_{S_A} |0\rangle_{L_A} , \qquad |-i\rangle_{S_A} |r\rangle_{L_A} \overset{U_A}{\longmapsto} |-i\rangle_{S_A} |1\rangle_{L_A} \tag{3.2}$$

and similarly for Bob and Charlie where the states $|0\rangle_{L_A}, |1\rangle_{L_B}$ denote the states of Alice's lab in which she obtains outcomes 0 and 1, respectively. This leads to the overall state

$$\left|\psi^{t=2}\right\rangle_{S_A L_A S_B L_B S_C L_C} = (U_A \otimes U_B \otimes U_C) \left|\psi^{t=1}\right\rangle_{S_A L_A S_B L_B S_C L_C} . \tag{3.3}$$

3. **Ursula, Valentina, and Wigner supermeasure.** Finally, the superobservers Ursula, Valentina, and Wigner respectively measure the systems $S_A L_A$, $S_B L_B$, and $S_C L_C$ in a basis with vectors[3]

$$|\text{yes}\rangle_{S_A L_A} = U_A |+\rangle_{S_A} |0\rangle_{L_A} , \qquad |\text{no}\rangle_{S_A L_A} = U_A |-\rangle_{S_A} |0\rangle_{L_A} . \tag{3.4}$$

We label the outcomes of $|\text{yes}\rangle, |\text{no}\rangle$ by $0, 1$.

[3]The two other basis vectors need not be specified since $|+i, 1\rangle_{S_A L_A}$ and $|-i, 0\rangle_{S_A L_A}$ do not occur.

The probabilities for Ursula's measurement in the $|\text{yes}\rangle, |\text{no}\rangle$-basis are equal to those which one would obtain by undoing Alice's measurement on S_A, followed by an X-measurement on the qubit S_A, and similarly for Valentina's and Wigner's measurements. The probabilities for $|\text{yes}\rangle, |\text{no}\rangle$ measurements by superobservers thus correspond to probabilities for X-measurements on the original GHZ state. For example, the unitary description of Alice's measurement followed by Ursula's measurement effect yields the probabilities

$$\left|\langle\pm|_{S_A}\langle 0|_{L_A} U_A^\dagger U_A |\ldots\rangle_{S_A}|0\rangle_{L_A}\right|^2 = \left|\langle\pm|_{S_A}\langle 0|_{L_A}|\ldots\rangle_{S_A}|0\rangle_{L_A}\right|^2 = \left|\langle\pm|_{S_A}|\ldots\rangle_{S_A}\right|^2. \tag{3.5}$$

Possible lightcone structures of this scenario are sketched in Fig. 3.2. Our results are independent of the exact causal structure, but we consider the natural causal structure arising from relativistic spacetime. More generally, we could embed the GHZ–FR scenario into a process-theoretic framework like Categorical Quantum Mechanics [184, 281] or the process operator formalism [282].

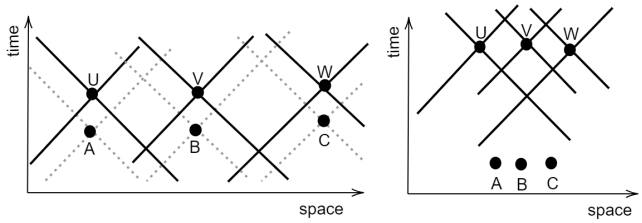

Figure 3.2: Possible lightcone structures of the agents in the GHZ–FR scenario, with Ursula, Valentina, and Wigner being superobservers for, respectively, Alice, Bob, and Charlie, who share a GHZ state. The GHZ–FR paradox occurs in both lightcone structures. A mixture of these lightcone structures is also possible.

Reasoning steps. Using the possibilistic Born rule, Ursula reasons about her outcome together with Bob's and Charlie's outcomes, and concludes: "*Having obtained u, I know that $b \oplus c = 1 \oplus u$*", since jointly considering the outcomes of Ursula, Bob and Charlie corresponds to the context XYY in the GHZ–Mermin model. For example, if Ursula obtained $u = 0$ then she finds that $b = c = 0$ is excluded as:

$$p(u = 0, b = 0, c = 0) = \mathrm{Tr}\,|\Psi\rangle\langle\Psi| = 0, \tag{3.6}$$

where

$$|\Psi\rangle = \langle+|_{S_A}\langle 0|_{L_A} U_A^\dagger U_A \langle+i|_{S_B}\langle+i|_{S_C}\left|\psi^{GHZ}\right\rangle_{S_A S_B S_C}|000\rangle_{L_A L_B L_C}\,. \tag{3.7}$$

Valentina and Wigner make similar reasoning statements, leading to

$$b \oplus c = 1 \oplus u, \quad a \oplus c = 1 \oplus v, \quad a \oplus b = 1 \oplus w. \tag{3.8}$$

with $\oplus$ meaning summation modulo 2. Applying the possibilistic Born rule to the superobservers' measurements with classically recorded outcomes, one obtains

$$u \oplus v \oplus w = 0. \tag{3.9}$$

For instance, we have

$$p(u = 0, v = 0, w = 1) = \mathrm{Tr}\, |\Psi_{u=0,v=0,w=1}\rangle \langle \Psi_{u=0,v=0,w=1}| = 0, \tag{3.10}$$

where

$$|\Psi_{u=0,v=0,w=1}\rangle = \langle +|_{S_A} \langle 0|_{L_A} \langle +|_{S_B} \langle 0|_{L_B} \langle -|_{S_C} \langle 0|_{L_C} U_A^\dagger U_A U_B^\dagger U_B U_C^\dagger U_C \left|\psi^{GHZ}\right\rangle_{S_A S_B S_C} |000\rangle_{L_A L_B L_C}. \tag{3.11}$$

Namely, this is the state obtained when Alice, Bob and Charlie perform their Y-measurements on their systems S_A, S_B, S_C, modeled unitarily by Ursula, Valentina and Wigner, who next undo these measurements and perform X-measurements with outcomes $0, 0, 1$, respectively. The proofs for other combinations of values for u, v, w with $u \oplus v \oplus w$ odd are entirely analogous. As Ursula, Valentina and Wigner come together to share their predictions, they find Eqs. (3.8) and (3.9). These equations correspond to Eq. (3.1) from the GHZ–Mermin scenario, with $x_A, x_B, x_C, y_A, y_B, y_C$ substituted for u, v, w, a, b, c. Namely, summing all of the equations of Eqs. (3.8) and (3.9) modulo 2 one finds $0 = 1$. Thus they find a contradiction.[4]

3.2 Two GHZ–FR no-go theorems

Having described the GHZ–FR paradox, we spell out the required assumptions in detail that lead to a contradiction. We construct two no-go theorems of increasing strength,

[4] Note that post-selected on u, v, w, the GHZ–FR argument gives a Specker's triangle [251]. For instance, for $u = v = w = 0$ we find

$$b \oplus c = 1, \quad a \oplus c = 1, \quad a \oplus b = 1. \tag{3.12}$$

which we name the GHZ–FR *truth* and *agreement* no-go theorems. The truth no-go theorem is easier to grasp as its assumptions are more natural. Moreover, it connects to a wider body of literature as it involves the assumption of Absoluteness of Observed Events, stating that performed measurements have absolute, single-valued outcomes, which is often invoked in extended Wigner's friend paradoxes [12–15, 20, 21, 23, 25–32, 266]. The agreement no-go theorem replaces Absoluteness of Observed Events by a weaker notion of agreement among classical agents. The fact that the paradox remains points towards a new principle to resolve extended Wigner's friend paradoxes. For each assumption, we state the general principle and specify its application to the GHZ–FR paradox.

3.2.1 GHZ–FR truth no-go theorem

Assumptions

The name 'the GHZ–FR *truth* no-go theorem' refers to the assumption of an underlying 'absolute truth' about the observations made in each *single run* of the protocol, formalised as Absoluteness of Observed Events (assumption A1) below [12–14].

We begin by presenting a list of assumptions that capture how we usually think about outcomes of performed measurements in experiments, namely the conjunction of possibilistic Born rule, Absoluteness of Observed Events, and Born Compatibility. Once we have established this common ground, we add the assumption that superobservers exist and that they can be described as in Section 2.3.1, Eq. (2.21), formalised as the assumption of Universality of Quantum Theory.

In quantum theory, the Born rule is used to obtain probability distributions for measurement outcomes. In situations that involve multiple observers and superobservers, one may expect the Born rule not to be applicable to all combinations of (actually performed) measurements. For example, it would be too much to expect it to apply to the measurements of Wigner and his friend, as Wigner may even undo the friend's measurement. In the present context however, only a minimal use of the Born rule is needed. We only require the Born rule to hold in situations in which agents believe they could still test their predictions. For example, in the GHZ–FR protocol Ursula makes predictions for Bob and Charlie, believing that she could check their outcomes, i.e. that she could still bring all relevant outcomes together. Indeed, for all she knows, Valentina and Wigner might not have performed their measurements yet; in fact, Ursula need not know even about their existence and thinks she can still obtain the outcomes of Bob

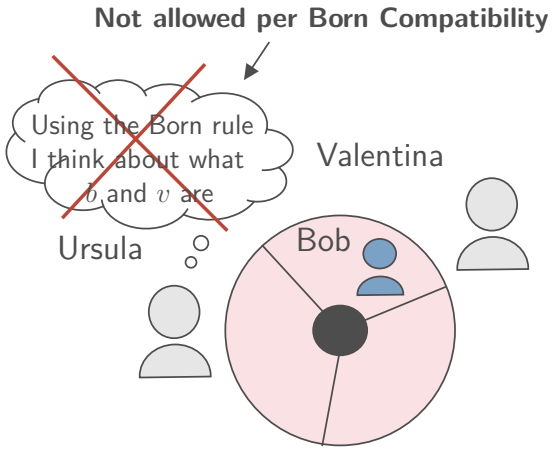

Figure 3.3: An example of a reasoning situation following the GHZ–FR protocol as in Fig. 3.1 that is not allowed using the possibilistic Born rule: Valentina is a superobserver for Bob, performing a supermeasurement on Bob's lab. Ursula cannot use the Born rule to make a single prediction involving both Bob's and Valentina's outcomes b and v, as she can never store both simultaneously in a memory.

and Charlie, making it natural for her to formulate a prediction about b, c. By contrast, one *cannot* apply the Born rule to the joint outcomes of measurements performed by a superobserver-observer pair, for example to Valentina's and Bob's measurements, in a single use of the rule as nobody could ever reasonably believe to obtain such outcomes jointly; see also Fig. 3.3. The Born rule may be valid more broadly, but we only require it be applicable in these instances.

Moreover, we only need *possibilistic* use of the Born rule.

The simplest requirement regarding the possibilistic Born rule is to require its experimental validity, i.e. that a classical agent can never experimentally disprove the possibilistic Born rule by gathering outcomes of measurements, possibly performed by other agents. We will refer to this principle as the possibilistic Born rule. We assume that the agent using the Born rule has, of course, a correct description of the measurements and states involved.

The principle possibilistic Born rule allows an agent to make predictions about joint measurement outcomes which are indeed borne out if they gather the outcomes of all involved measurements. The stronger requirement that such predictions made by an agent be *valid* even if the agent does not actually bring those measurement outcomes together can be phrased as follows:

> Every performed measurement has an absolute, single-valued outcome about which the possibilistic Born rule as used by agents provides valid predictions.

We can split this principle up into two philosophical principles: the existence of an absolute truth, formalised in Absoluteness of Observed Events, and possibilistic Born rule as used by agents vis-à-vis this absolute truth, regardless of whether its predictions are verified by gathering all involved outcomes, formalised in Born Compatibility.[5]

Assumption A1 (Absoluteness of Observed Events (AOE))**.** Every performed measurement has an absolute, single-valued outcome.

Concretely, in each run of the protocol, there exists an assignment $f : \mathcal{M} \to O$, where $\mathcal{M}$ is the (finite) set of all the measurements performed by agents in the protocol, mapping each performed measurement to its observed outcome.[6]

We now require the absolute outcomes of AOE to be compatible with an agent's use of the Born rule, making the agent's statements obtained from their application of the Born rule true regardless of whether the agent actually brings the involved outcomes together, but only when the agent believes to be able to gather the relevant outcomes in a joint memory.

Assumption A2 (Born Compatibility)**.** Outcomes assigned per AOE to a set of measurements that an agent A believes she can gather jointly in a memory must be compatible with A's use of the possibilistic Born rule.
Concretely, the assignment $f : \mathcal{M} \to O$ per AOE is compatible with all allowed uses of the possibilistic Born rule by agents.

Using Absoluteness of Observed Events and Born Compatibility an agent may translate their predictions from possibilistic Born rule to conditions on $f : \mathcal{M} \to O$. For example, if the Born rule deems impossible (i.e. assigns probability zero to) the joint outcome $(a_1, \ldots, a_n)$ of the measurements $M_1, \ldots, M_n$, then in each run of the protocol the 'observed outcome' assignment f cannot satisfy $f(M_i) = a_i$ for all $i = 1, \ldots, n$.

Without any further restriction, the AOE assumption is mathematically trivial, i.e. a value assignment $f : \mathcal{M} \to O$ can always be found in each round of a measurement protocol such as the GHZ–FR. However, Born Compatibility introduces restrictions on

[5] In epistemic logic, Absoluteness of Observed Events could be formalised as the Truth axiom [23, 30, 31]; see Remark 3 ahead for the formulation of key assumptions in our no-go results in the language of epistemic modal logic.

[6] For simplicity, we take all measurements to be valued on the same outcome set O. If each measurement m has its own set O_M of (a priori) possible outcomes, then the type of f would more properly be that of a dependent function, or a tuple, $f \in \prod_{m \in \mathcal{M}} O_m$, but we might as well take $O = \cup_{m \in \mathcal{M}} O_m$.

such value assignments, forcing them to conform with testable predictions, thereby rendering the underlying absoluteness requirement non-trivial.

Remark 1 (AOE and noncontextuality). The assumption Absoluteness of Observed Events closely resembles that of noncontextuality in the sheaf-theoretic framework [42]. In FR-like paradoxes the different contexts required for a contradiction are operationally brought together through the statements made by different agents. Physically, the difference with respect to Bell nonlocality scenarios lies in the fact that in FR-like paradoxes all the measurements that are assigned outcomes are actually performed in each single round of the protocol. The assumptions Absoluteness of Observed Events and Born Compatibility do not yield contradictions in quantum experiments without superobservers,[7] such as standard Bell or contextuality scenarios. In such setups, after each run of the experiment there exists a consistent assignment of outcome values to all the measurements that *have been performed* in that run. Therefore, Absoluteness of Observed Events does not involve any counterfactual statements about the outcomes of unperformed measurements, as in Bell locality or noncontextuality assumptions, since all involved measurements are actually performed by some agent. The only 'counterfactual' aspect here is captured by the fact that the possibilistic Born rule is required to be valid regardless of whether the outcomes of the performed measurements are actually brought together, as per Born Compatibility.

The two assumptions A1 and A2 capture how we normally reason in (quantum) experiments. Our final assumption introduces superobservers.

Assumption A3 (Universality of Quantum Theory)**.** An observer can perform a measurement on a quantum system. Furthermore, a sealed lab in which an observer performs a measurement on some system can be described as a closed quantum system that evolves unitarily, upon which a superobserver can apply quantum operations, including measurements.

Concretely, this assumption ensures that the GHZ–FR protocol including superobservers (pictured in circuit form in Fig. 3.1, with the parties situated as in Fig. 3.2) can be performed. More specifically: first, there exists a GHZ state, upon which Alice, Bob, and Charlie can perform their Y-measurements; these measurements and everything else happening in each of the observers' sealed labs can be modeled as a known unitary by a su-

[7] Even with a single observer-superobserver pair as in Wigner's original thought experiment, these two assumptions are not contradictory, as after each single run of the experiment there always exists a consistent history of outcome assignments.

perobserver; secondly, superobservers can perform supermeasurements on observers, that is, Ursula, Valentina, and Wigner can perform the X-like supermeasurements described in Section 3.1.2.

We are now in a position to state the GHZ–FR truth no-go theorem.

Theorem 2 (GHZ–FR truth no-go theorem)**.** *The GHZ–FR paradox shows that the Universality of Quantum Theory, Absoluteness of Observed Events and Born Compatibility are incompatible.*

Proof. The assumption Universality of Quantum Theory ensures that the GHZ–FR measurement protocol can be performed. For the reasoning stage of the paradox, the following assumptions are required. In each run of the protocol, by Absoluteness of Observed Events, there exists a function $f\colon \mathcal{M} \to O$ assigning an outcome to each performed measurement. Writing A, B, C, U, V, W for the measurements performed in the GHZ–FR paradox, such assignment f consists of outcome values $f(A), f(B), f(C), f(U), f(V), f(W)$, which were denoted by a, b, c, u, v, w, respectively, in Section 3.1.2.

Using Born Compatibility (which can be applied since Ursula has no knowledge about Valentina and Wigner and thus believes she is able to gather u, b, c jointly), Ursula reasons that this assignment must satisfy

$$f(B) \oplus f(C) = 1 \oplus f(U), \tag{3.13}$$

and similarly Valentina and Wigner respectively reason that

$$f(A) \oplus f(C) = 1 \oplus f(V), \tag{3.14}$$
$$f(A) \oplus f(B) = 1 \oplus f(W). \tag{3.15}$$

Finally, applying possibilistic Born rule (implied by Born Compatibility) to the classically recorded outcomes of U, V, W, which are available after the protocol, we know that these must satisfy

$$f(U) \oplus f(V) \oplus f(W) = 0. \tag{3.16}$$

But no assignment f can simultaneously satisfy the four required equations, leading to a contradiction. ■

Ways to resolve the truth no-go theorem

The assumptions of the GHZ–FR truth no-go theorem (Theorem 2) are shown in Fig. 3.4. The theorem forces us to reject at least one of these three assumptions. We discuss each of the three possibilities in turn.

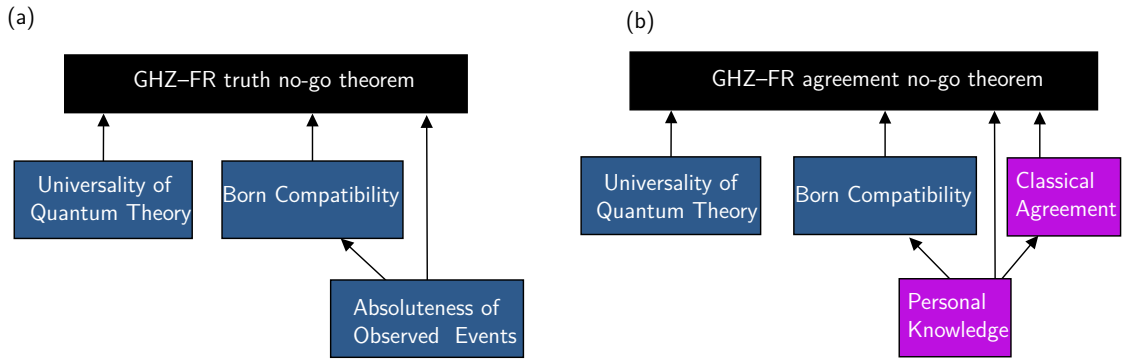

Figure 3.4: Assumptions leading to (a) the GHZ–FR truth no-go theorem of Section 3.2.1 and (b) the GHZ–FR agreement no-go theorem of Section 3.2.2, where AOE is replaced by Personal Knowledge and Classical Agreement. An arrow from one assumption box to another denotes that the former is used to formulate the latter.

a. Reject Absoluteness of Observed Events Refuting Absoluteness of Observed Events means one cannot simply talk about single-valued outcomes obtained in measurements performed by other agents as being absolute, i.e. a function $f : \mathcal{M} \to O$ assigning an outcome to each performed measurement cannot exist. Still, we would like agents to be able to formulate predictions correctly. Therefore, we may ask 'When can an agent assign a single-valued outcome to a performed measurement?' Naturally, the experimenter who performs the measurement can assign a single-valued outcome, namely the one they observe. This extends also to any agent who learns about the outcome, i.e. any agent obtaining a copy of the outcome record and storing it in their memory[8]. Can other agents assign an outcome – that is not a priori absolute – too? We probe this possibility in the GHZ–FR agreement no-go theorem in the next section, where we weaken Absoluteness of Observed Events to allow for each agent to assign a personal single-valued outcome to performed measurements, which may a priori differ among different agents. The stronger no-go result obtained indicates that this approach may not lead to a resolution, and that perhaps only the experimenter and agents learning about the outcome can assign an outcome value to a performed experiment.

[8]For a more elaborate definition of *learning about an outcome*, see Section 3.2.3 below.

b. Reject Born Compatibility As Born Compatibility assumes AOE, refuting AOE implies refuting Born Compatibility. However, one may also consider accepting AOE but refuting Born Compatibility. In that case, there is still an absolute single-valued outcome assigned to each performed measurement, but the use of the Born rule per possibilistic Born rule does not provide valid predictions about such assignment. In other words, more cautious use of the Born rule is required. Therefore, one needs an additional prescription for how all agents can use the Born rule. This is the route typically taken by (contextual) hidden variable theories. In this case, the observers' memories may also be described by hidden variables, such as in Bohmian theories [283–285], making correct use of the Born rule more subtle. We will discuss this option further in Section 3.2.2.

When refuting Born Compatibility, one can instead go further as well and refute the experimental validity of the Born rule (possibilistic Born rule) as well, i.e. modify the Born rule for all applications rather than only the non-experimental ones.

c. Reject Universality of Quantum Theory To reject the existence of superobservers, however, comes with its own complications. If we adopt the viewpoint that only classical observers can perform measurements, where do we draw the line? Would a hypothetical quantum computer showing signs of artificial intelligence not qualify as an observer? Or would the notion of a classical observer arise as an emergent concept? Alternatively, one may reject the notion of measurement altogether.

In the foundations of quantum theory, the above questions deal with where to place the Heisenberg cut – or whether it even exists. In fact, Wigner's original thought experiment probed precisely this question. However, with the FR and GHZ–FR paradoxes, the problem gains new importance, due to the stronger implications of accepting superobservers. Namely, seemingly innocent use of quantum theory leads to contradictory statements regarding not only the question of *when* a measurement does happen and classical outcomes are produced, but also regarding *what* actual outcome values were observed. In short, inconsistent histories may be obtained by different observers.

In other words, if one accepts the experimental validity of the (possibilistic) Born rule and the existence of superobservers A3, the contradiction in the GHZ–FR paradox arises through the assumed existence of a function $f\colon \mathcal{M} \to O$ assigning absolute, single-valued outcomes to all performed measurements which are compatible with the agents' allowed uses of the Born rule. In that case, one must refute either the existence of such an absolute assignment f, thus rejecting AOE A1, or the compatibility of f with the agents'

use of the Born rule, rejecting Born Compatibility A2 and require more cautious use of the Born rule.

We refer to Section 3.3 for a discussion of the FR assumptions, responses in literature and their relevance for the GHZ–FR no-go theorems.

3.2.2 GHZ–FR agreement no-go theorem

Assumptions

We now replace Absoluteness of Observed Events by a weaker condition, split up into two assumptions. We introduce Personal Knowledge, which a priori allows different agents to assign different single-valued outcomes to the same measurement, but subject it to a consistency condition called Classical Agreement, which stipulates that a statement known to be true by a classical agent can be taken to be true by any other classical agent to whom it is communicated. As in the truth no-go theorem, we require the outcome assignments per Personal Knowledge to be compatible with the Born rule. These assumptions allow superobservers Ursula, Valentina and Wigner to formulate their predictions about outcomes and share them with each other, obtaining a contradiction.

We stress that only classical agents need to reason and communicate; there is no need for classical agents to trust the reasoning of quantum agents as in Ref. [11, 30, 31]; see also Remark 3 below.

Assumption A1′ (Personal Knowledge)**.** Every performed measurement can be assigned a single-valued outcome by an agent who believes they are able to gather that outcome.[9] Concretely, in each run of the protocol, for each agent A, there exists an assignment $f_A : \mathcal{M}_A \to O$, where $\mathcal{M}_A$ is the set of performed measurements whose outcomes could be gathered from A's perspective, mapping each performed measurement to an outcome value assigned by A.

The agents (superobservers) assigning outcomes in the GHZ–FR paradox only do so *after* having performed their own measurement, i.e. based on what they believe is available to them then.

We recall that Born Compatibility builds on AOE, but we can also state an analogous Born rule compatibility requirement for this weakened version of AOE.

[9] For example, in the GHZ–FR paradox Ursula applies Personal Knowledge to assign values to Bob's and Charlie's measurements, whose outcomes she believes she could still obtain as she does not need to know about Valentina's or Wigner's existence or operations.

Assumption A2′(i) (Born Compatibility)**.** Outcomes assigned *by an agent* A per Personal Knowledge to a set of measurements must be compatible with A's use of the possibilistic Born rule.
Concretely, for every agent A, the assignment $f_A \colon \mathcal{M}_A \to O$ per Personal Knowledge is compatible with all allowed uses of the possibilistic Born rule by A.

In Personal Knowledge, each function f_A assigns outcomes to measurements, but, crucially, these outcomes need not be absolute but can be relative to agent A. Using Born Compatibility, agent A may translate her predictions from possibilistic Born rule to conditions on $f_A : \mathcal{M}_A \to O$. For example, if the Born rule deems impossible the joint outcome $(a_1, \ldots, a_n)$ of the measurements $M_1, \ldots, M_n$, then the assignment f_A cannot satisfy $f_A(M_i) = a_i$ for all $i = 1, \ldots, n$. Agent A may conclude:

> 'I (agent A) know that there exists a measurement $M_i \in \{M_1, \ldots, M_n\}$, possibly performed by other agents, that does not have outcome a_i (for me), regardless of whether I ask the involved agent(s) for it.

Next, we state a minimal practical requirement on the outcome assignments across different classical agents. This assumption was not needed for the GHZ–FR truth no-go theorem, as it is implied by AOE and Born Compatibility.

Assumption A2′(ii) (Classical Agreement)**.** The outcome assignments (per Personal Knowledge) of classical agents who communicate classically must agree on their overlap. Concretely, if classical agents A and B classically communicate their outcome assignments, then these assignments f_A and f_B must agree on their overlap $\mathcal{M}_A \cap \mathcal{M}_B$.

In fact, the way we use this assumption is that, for example, Ursula and Valentina, must agree on their outcome assignment for the measurement C: $f_U(C) = f_V(C)$. In fact, in a version of the protocol where Wigner performs his measurement later, both Ursula and Valentina can ask Charlie for his outcome value. They will obtain the same value, motivating this assumption. In our particular protocol, Personal Knowledge and Classical Agreement will lead to agents U, V, W having to agree on their assignments so that again a single assignment f to all performed measurements can be built that must satisfy the GHZ–Mermin equations.

Theorem 3 (GHZ–FR agreement no-go theorem)**.** *The GHZ–FR paradox shows that the assumptions Universality of Quantum Theory, Personal Knowledge, Classical Agreement, and Born Compatibility are incompatible.*

Proof. The first part of the proof is analogous to the proof of Theorem 2, replacing the absolute outcome assignment f by the outcome assignments f_U, f_V, f_W of Ursula, Valentina and Wigner per Personal Knowledge. For example, as Ursula has no knowledge about Valentina and Wigner and thus believes she is able to gather u, b, c jointly, she can assign them outcome values $f_U(B), f_U(C)$ (and $f_U(U)$ of course), compatible with her Born rule use per Personal Knowledge and Born Compatibility. Then Classical Agreement requires these outcome assignments to agree on their overlaps as Ursula, Valentina and Wigner are treated classically in the protocol and can physically meet and communicate as much as they like, so that a single global outcome assignment can be built that satisfies the GHZ–Mermin equations. Details can be found in Appendix B of Ref. [19]. ■

Remark 2 (Heisenberg cuts). In literature, sometimes A applying the Born rule to set of jointly performed measurements is also referred to as A choosing a Heisenberg cut, having the performed measurements outside the cut. In that terminology, our assumptions of Personal Knowledge and Born Compatibility can also be phrased as: 'An agent A can choose a Heisenberg cut, placing a set of measurements she thinks she can still gather jointly all outside the cut.' The paradox then arises if it is understood that different choices of Heisenberg cuts, when chosen by classical agents, must not lead to contradictions about the actual outcome values, which is another way of phrasing Classical Agreement. Interpretations of quantum mechanics which resolve the measurement problem by determining or eliminating the Heisenberg cut will resolve EWF arguments by providing an unambiguous description for the entire protocol. We discuss Heisenberg cuts in more detail in Section 3.4.

Remark 3 (The GHZ–FR assumptions in modal logic). We briefly outline a discussion of the GHZ–FR paradox in terms of epistemic logic, and refer to Appendix C of Ref. [19].

Syntactically, in epistemic logic [286, 287], besides the usual logical connectives, formulas can be built using a unary modal operator K_A for each agent A, with a formula $K_A\phi$ interpreted as 'agent A believes (or knows) that ϕ is True'. Different sets of axioms are typically considered in epistemic logic systems, varying with the intended interpretation.

In Ref. [30], the *Distribution Axiom* of modal logic (a.k.a. axiom **K**) is listed as assumption (D) of the FR paradox (see Section 3.3), and used to justify the use of *modus ponens* by quantum agents in the FR paradox. In the GHZ–FR paradox, however, as in the version of the FR paradox from Section 2.4.2, only classical agents need to reason, dispensing with the reasoning by quantum agents considered in the earlier treatments

of the FR paradox [11, 23, 30, 31]. The fact that we need only consider reasoning by classical agents in the protocol is the reason why we do not state such reasoning as a separate assumption in our no-go theorems. In modal logical terms, this is reflected in the fact that we do not need the epistemic modalities K_A for quantum agents: the whole argument can be phrased in a logical system with modalities only for each *classical* agent in the protocol. It could perhaps be argued that this somewhat weakens the appeal of a formalisation of the paradoxes using modal logic.

Still, the core assumptions of our GHZ–FR no-go theorems can be phrased in these terms. The existence of an absolute truth, per Absoluteness of Observed Events, in the GHZ–FR truth no-go theorem corresponds to the Truth axiom: if an agent *knows* something, then it is true: $K_A\phi \to \phi$. The agreement no-go theorem requires a form of agreement only among classical agents. It can be seen as a considerable weakening of the epistemic trust structures identified for the FR paradox in [23, 30, 31], in that we do not need classical agents to believe the reasoning of quantum agents.

Ways to resolve the agreement no-go theorem

The truth and agreement no-go theorems differ only in that the AOE assumption in the truth no-go theorem is replaced by Personal Knowledge and Classical Agreement in the agreement no-go theorem (cf. Fig. 3.4). If one opts to resolve the truth no-go theorem by rejecting AOE, then one must reject one of the assumptions Personal Knowledge and Classical Agreement to resolve the agreement no-go theorem. We argue that, in that case, the most practical resolution is to further weaken Personal Knowledge. Other resolutions are of course possible.

We now discuss the repercussions of separately rejecting each of the assumptions of the agreement no-go theorem. The consequences of rejecting Universality of Quantum Theory are the same as in the truth no-go theorem, as discussed in Section 3.2.1.

a. Reject Personal Knowledge Rejecting Personal Knowledge implies that one also rejects Absoluteness of Observed Events of the truth no-go theorem. If the outcome of a performed measurement is not absolute, one must ask: 'Who can assign a single-valued outcome to a performed measurement?'[10] Certainly an experimenter can assign a single value to the outcome of a measurement performed by themself and test the Born rule in this way – but who else?

[10] As the outcome of a performed measurement is considered to be classical information, a physical entity assigning an outcome to a measurement needs a preferred basis in its quantum storage system, as argued in Refs. [267, 268], which constitutes our definition of an agent.

Suppose that an agent A's prediction about outcomes of performed measurements is valid 'no matter whether A actually brings the involved outcomes together'. Then, one is assuming the existence of absolute outcomes or outcomes relative to the agent A as in Personal Knowledge. Therefore, rejecting AOE and Personal Knowledge suggests that only the experimenter themself can assign a single-valued outcome to a performed measurement, and by extension any other agent learning about the outcome. This proposal suggests a new principle: Relativity of Observed Events[11]:

> Relativity of Observed Events Every performed measurement can be assigned a single-valued outcome by an agent who learns about that outcome.

Consequently, an agent A can reason about the outcome b of a measurement performed by another agent B only to make statements about the value of this outcome b *conditioned on* agent A asking B for it, formalised in the following principle:

> Born Practicality An agent A can translate predictions based on the Born rule only into statements about outcomes that A would obtain when A learns about these outcomes.

To *learn about an outcome*, an agent performs an operation such that they, directly or indirectly, get to know the outcome result of the measurement through interactions with the experimenter (or other agents who know the result), measured system or environment, namely by (i) directly asking the experimenter, (ii) measuring the measured system in the same measurement basis or (iii) indirectly obtaining the measurement outcome through traces in the environment (which can also include being told about it by another agent).

b. Reject Born Compatibility If one rejects Born Compatibility, one can still keep AOE (and Personal Knowledge). However, in that case, as the GHZ–FR paradox shows, one needs a rule for when to expect Born rule predictions to be valid, as agents need to be able to phrase their Born rule predictions correctly. When holding on to the experimental validity of the Born rule per possibilistic Born rule, as a principle for when such predictions are valid, one can use Born Practicality as well. If an agent has more knowledge of the underlying theory or ontology their world is subject to, they can perhaps make more precise statements. For example, one can imagine a theory or ontology where AOE is satisfied (as in Bohmian mechanics), and where an agent can say their prediction is valid as long as they can gather all involved outcomes. However, one can imagine a

[11]The seed of this principle originated in a discussion of L.W. with E. Cavalcanti about the implications of the Local Friendliness no-go theorem [12, 13], but it is not necessarily what E. G. Cavalcanti believes to be a resolution.

theory where additionally actually asking for the outcomes changes the actual outcomes through hidden variables. Therefore, without further knowledge, in theories rejecting Born Compatibility, agents can use Born Practicality for correct predictions.

c. Reject Classical Agreement Rejecting Classical Agreement would mean that classically communicating agents (communicating in whichever way they prefer) disagree on the truth value of a statement about what happened in a measurement. In principle, every performed measurement whose outcome is not stored in some classical agent's memory can be part of a paradox à la GHZ–FR. Thus, disagreement among classically communicating agents about what happened in the measurement M topples a basic assumption of science: we cannot trust other scientists' reports of what happened in an experiment. To see this difficulty in more detail, we consider the following situation. Alice performs a measurement, and Bob makes a statement about Alice's outcome, claiming for example that Alice obtained $a = 0$. Bob could have arrived at this statement by reasoning or by explicitly asking Alice. Holding on to Personal Knowledge and Born Compatibility, there is no need for Bob to specify how he obtained the statement, which is true *for him*. Therefore, when Bob tells us that Alice obtained $a = 0$, we simply cannot know whether his statement (or any statement derived from it) is true *for us*. A potential way out is that Bob conveys more details about how he obtained his conclusion, and then we decide whether or not to believe him, as argued in QBism [288], for example. In our everyday world, such disagreement between agents does not arise, because of the absence of superobservers.

3.2.3 A possible resolution

In Section 3.2.2 we proposed Born Practicality to resolve the GHZ–FR paradox, and by extension all extended Wigner's friend paradoxes, if one accepts the existence of superobservers. The principle says that observers should state their Born rule predictions *conditioned* on learning about the involved outcomes. In theories that reject AOE, this principle can be implemented as Relativity of Observed Events, as explained in Section 3.2.2. This resolution can be seen as an agent-based extension of Peres's dictum that 'Unperformed experiments have no results' [39],

> Agent-based Peres's dictum: 'Unperformed experiments have no results, and unknown results have no values.'

We can see this as an extension of Peres's dictum, by considering learning about the outcome result (of a measurement performed by another agent) as an experiment too. Namely, you cannot *a priori* assign outcome values to measurements or experiments, possibly performed by other agents, you can only do so upon actually learning the outcome. In a sense, this principle is a natural classical vs. quantum separation. If a superobserver Wigner can model his friend's measurement unitarily, i.e. as a quantum system, then Wigner cannot assign classical information to his friend's measurement outcome, as this information is still quantum for Wigner. Instead, Wigner should only assign a single value to his friend's outcome when he learns about that outcome. Here is an example of Born Practicality in use for a simple Wigner's friend scenario:

> Wigner finds that $f = 1$ is impossible (has probability zero) with f denoting his friend's outcome and may conclude: "Using the Born rule I (Wigner) know that my friend would never reply that she obtained $f = 1$ *if* I were to ask her for her outcome, (as long as my friend's memory record and mine are not wiped or updated)."

Definition of 'learning about an outcome'. By 'learning about an outcome of a measurement', or 'asking an experimenter for the outcome of their performed measurement', we mean the following. Consider an agent A performing a measurement on a qubit system S in the basis $|0\rangle, |1\rangle$. Let E be the environment, namely the whole closed system excluding A and S. We can model this measurement as a unitary:

$$\Big(\alpha\,|0\rangle+\beta\,|1\rangle\Big)_S\otimes|\text{ready}\rangle_A\otimes|\text{ready}\rangle_E \to \alpha\,|0\rangle_S\otimes|0\rangle_A\otimes|0\rangle_E+\beta\,|1\rangle_S\otimes|1\rangle_A\otimes|1\rangle_E \quad (3.17)$$

where we assume that the agent and environment are initialized in the $|\text{ready}\rangle_A, |\text{ready}\rangle_E$. Here $|0\rangle_E, |1\rangle_E$ denote the traces of the measurement outcome 0 or 1 in the environment. Learning about the outcome of this measurement then includes:

- Asking the experimenter A for the outcome, corresponding to measuring A in the basis $|0\rangle_A, |1\rangle_A$;
- Measuring the system S in the measurement basis $|0\rangle_S, |1\rangle_S$;
- Obtaining the measurement outcome through traces in the environment, corresponding to measuring the environment E in the basis $|0\rangle_E, |1\rangle_E$;
- classical communication about outcomes through another agent who learns about the outcome in one of the three ways stated above.

Resolving the paradox. Let us briefly explain how assuming Born Practicality instead of AOE (or instead of Personal Knowledge) and Born Compatibility resolves the GHZ–FR paradox. The FR paradox is resolved analogously.

We assume that the GHZ–FR measurement protocol has been performed, considering a run with the (classically available) outcomes $u = 0, v = 0, w = 0$, say. Ursula, making her predictions, finds using Born Practicality:

> Ursula: 'if I (Ursula) were to ask Bob and Charlie for their outcomes b, c, then I would find that $b \oplus c = 1$, as long as my memory and theirs are not wiped or updated.'

Bob and Charlie argue similarly to conclude that $a \oplus c = 1$ and $a \oplus b = 1$ when they ask for (a, c) and (a, b), respectively. However, none of these superobservers have actually asked the other observers Alice, Bob and Charlie for their outcomes, so that no paradox arises. All precise statements, as well as those for the FR paradox, can be found in Appendix D of Ref. [19].

3.3 Comparison with the FR paradox

We compare the GHZ–FR paradox to the FR paradox in Section 3.3.1. Then, in Section 3.4, we place our work within the broader literature on the subject: we briefly review previous responses to the FR paradox, as well as some work closely related to the GHZ–FR paradox, where similar measurement protocols have been proposed but with different analyses.

3.3.1 Why the GHZ–FR paradox is stronger than the FR paradox

To establish the GHZ–FR paradox, in stark contrast to the FR paradox [11, 24], (i) only *classical* agents reason, and (ii) *no post-selection* on outcomes is required. The Born rule is only used to determine that certain (joint) measurement outcomes are *impossible*. It need not be used to assert that certain outcomes are *possible*, that is, that they occur with (some) nonzero probability. While the former are definitive statements about *every* run of the experiment – they state that a particular combination of outcomes is *never* witnessed –, the latter inform about events that occur only in *some* runs, which would need to be post-selected for the derived contradiction to apply. This difference is at the root of distinctive feature (ii) above.

To compare the assumptions of the FR no-go theorem and those of the GHZ–FR agreement no-go theorem, the strongest of the two no-go theorems from Section 3.2, we begin by listing the assumptions of both theorems (cf. Fig. 3.5). For the FR theorem, we rely on the presentation in Ref. [24], which considers the following list of assumptions: (C) Consistency, (D) Distribution axiom of modal logic, (Q) Quantum Born rule statements, (U) Unitarity, and (S) Single-valuedness of outcomes.

Distinctive feature (i) – that only classical agents reason – is reflected in the weakening of assumptions (C), (D), and (Q) in the GHZ–FR paradox. Distinctive feature (ii) – that no post-selection is required – is not reflected in the assumptions. It corresponds to the fact that the contradiction at the heart of the GHZ–FR paradox occurs in *every* single round of the protocol, whereas the FR paradox only obtains a contradiction in *some* (postselected) rounds in which a specific value of the final (super)measurements is observed.

The FR assumptions. We first summarize the assumptions of the FR no-go theorem in our own wording. All uses of the word 'statement' are shorthand for 'statement about outcome values observed by agents'.[12]

(Q) An agent A can translate the use of the possibilistic quantum Born rule applied to a set of performed (not necessarily all by A) measurements, to make statements about what outcomes are obtained or not.

(C) If an agent B knows that statement s_1 is true from reasoning with a theory that an agent A accepts[13] and A knows about this reasoning of B about s_1, then s_1 is true for agent A as well.

(D) If an agent A knows a statement s_1 and also knows that s_1 implies statement s_2 then agent A can conclude 'I (A) know s_2'. [14]

[12] In the FR, one refers to outcome values observed by a single agent in these statements. In the GHZ–FR multiple agents are involved in such statements, for example Ursula reasons about b, c observed by Bob and Charlie, respectively; Ursula treats these joint outcomes as an outcome of a joint measurement by Bob and Charlie. However, in the GHZ–FR no-go theorem we could also allow A, B, C to share their outcome values with each other so that both b, c are observed by Bob and Charlie for example, and have a superobserver later undo this sharing before other measurements are performed. We could even have all these measurements A, B, C be performed by the same agent, with U, V, W acting upon different parts of that agent's memory, so that when Ursula talks about outcomes b, c this refers to a joint measurement performed by that single agent.

[13] By a *theory that an agent A accepts*, we mean reasoning according to these five rules.

[14] Assumption (D) refers to the Distribution axiom of Modal Logic and was first identified in Ref. [30], along with assumption (U).

(S) Outcomes of performed measurements from an agent's viewpoint who uses (Q) are single-valued.

(U) An agent A can model another agent B measuring a system as a unitary process.

Let us give an example of (Q) in use. If agent A applies the Born rule to a set of measurements that can be simulated in parallel and finds that a certain outcome $(a_1, \ldots, a_n)$ cannot occur, then A can conclude that some a_i is not obtained (in a measurement performed possibly by another agent).[15]

Rephrasing the GHZ–FR assumptions We now rephrase the GHZ–FR assumptions in order to simplify the comparison with the assumptions of the FR paradox, *emphasizing* the differences.[16]

(Q') A *classical*[17] agent A can translate the use of the possibilistic quantum Born rule applied to a set of performed (not necessarily all by A) measurements, to make statements about what outcomes are obtained or not.

(C') If a *classical* agent B knows that statement s_1 is true from reasoning with a theory that a *classical* agent A accepts, and B *communicates his reasoning to* A *through classical communication* (so that A knows about this reasoning of B about s_1), then s_1 is true for agent A as well.

(D') If a *classical* agent A knows a statement s_1 and also knows that s_1 implies statement s_2 then agent A can conclude 'I (A) know s_2'.

Comparing FR and GHZ–FR. Comparing the FR and GHZ–FR assumptions, (U) and (S) are the same, but the GHZ–FR assumptions (C'), (D'), (Q') are weaker versions of the FR assumptions (C), (D), (Q), respectively (cf. Fig. 3.5). Namely, only *classical agents* need to produce statements and reason, and accept statements classically communicated among them.[18] This is an important weakening, as for example the Sleeping Beauty

[15] We rephrased assumption (Q) here somewhat compared to FR's original phrasing, as unwanted interpretations of the latter may (wrongly) conclude collapse to be necessary for the FR paradox [27, 272, 289–291].

[16] In fact, this paraphrasing of our no-go theorem has slightly stronger assumptions than needed. We did so to tailor them more to the phrasing of the FR assumptions. To weaken them, we could allow the (Q) assumption to have statements about outcome values assigned by the agent making the prediction, subject to Classical Agreement.

[17] We recall that by a *classical* agent we mean an agent that need not be modeled quantumly by any other agent in the protocol.

[18] As described in Section 2.4.2, a slight modification of the reasoning in the FR paradox can also produce a no-go theorem from these weaker assumptions. A similar modification of the reasoning in the FR paradox can be found in Ref. [270].

problem of Ref. [292] shows how FR's assumption (C) might be problematic already in a classical setup, undermining the motivation for this assumption (C) in addition to establishing that a violation of (C) is certainly not what is especially quantum about extended Wigner's friend paradoxes. Their argument does not apply to its weakened version (C') in the GHZ–FR paradox.

Explaining the rephrasing of the GHZ–FR assumptions. We have argued that the GHZ–FR paradox requires assumptions (U), (S), (C'), (D'), and (Q'), modeled after those used in discussions of the FR paradox. We now explain how they relate to the concepts used to describe the GHZ–FR agreement no-go theorem of Section 3.2.2; see also Fig. 3.5. Concerning the truth no-go theorem, we recall that the assumptions Classical Agreement and Personal Knowledge of the agreement no-go theorem are replaced by the *a priori* stronger Absoluteness of Observed Events in the truth no-go theorem.

(U) is the equivalent of the GHZ–FR assumption Universality of Quantum Theory.

(S) is captured in Personal Knowledge, and Absoluteness of Observed Events, allowing an agent to assign a single-valued outcome to performed measurements.

(C') is a (slightly stronger) paraphrasing of Classical Agreement.

(D') allows for classical agents to reason and is assumed implicitly in the GHZ–FR agreement no-go theorem. Classical agents are just like us, using elementary rules of logic such as the *modus ponens* as expressed by assumption (D').

(Q') is captured in the conjunction of AOE or Personal Knowledge and Born Compatibility. The assumption (Q') assumes the validity of the Born rule. But in fact, (Q') needs more, as it allows an agent A to make statements about outcomes obtained in a measurement performed by another agent B, regardless of whether A tests this Born rule prediction or not, captured by Personal Knowledge and Born Compatibility, respectively.

3.4 Related work on the FR paradox

In this section, we will compare the GHZ–FR paradox to previous work on the FR paradox and protocols related to the GHZ–FR paradox.

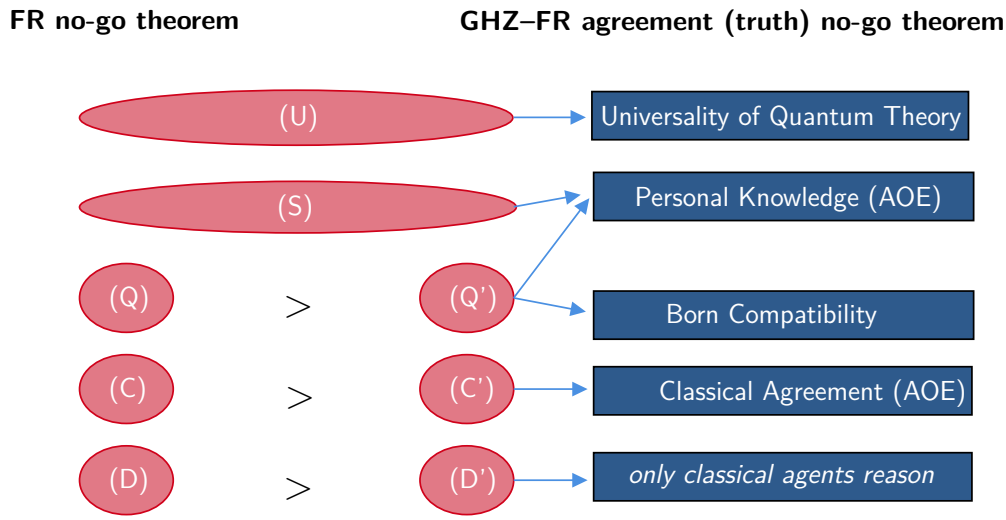

Figure 3.5: Comparing the assumptions of the FR paradox to those of the GHZ–FR no-go theorems. An assumption of the FR no-go theorem being strictly stronger than a corresponding assumption of the GHZ–FR agreement no-go theorem is denoted by $>$. In red we compare the assumptions (U), (S), (Q), (C) and (D) of the FR no-go theorem to the rephrased GHZ–FR assumptions (U), (S), (Q'), (C') and (D'). How these rephrased assumptions correspond to the GHZ–FR assumptions of Section 3.2, which we have used throughout this work, is shown by the arrows between the assumptions in red circles and blue boxes. The assumptions of the agreement no-go theorem are listed, as the agreement no-go theorem is stronger than the truth one. In the truth no-go theorem, Personal Knowledge and Classical Agreement are replaced by the stronger Absoluteness of Observed Events, as listed in brackets.

3.4.1 Previous responses to the FR paradox

The FR no-go theorem [11, 24] forces us to refute one of its assumptions, and has been discussed extensively in the literature. We are not able to survey all relevant works; instead, we aim to provide a concise overview. We classify previous responses to the FR paradox into three main categories.[19] For more extended reviews we point the reader to Refs. [24, 264].

First, there is work that questions and re-analyses the assumptions used in the FR paradox [23, 24, 27, 30, 31, 264, 285, 293, 294]. Some implicit assumptions have been identified in Refs. [30, 285]. A link between paradoxes in FR-like scenarios and multi-agent knowledge structures, epistemic logic [286, 287], and contextuality has been investigated in Refs. [23, 30, 31, 293–296].

Other works claim that the FR paradox contains implicit assumptions that can be deemed invalid a priori, thereby questioning the relevance or novelty of the FR paradox

[19] Some works belong to more than one category, so we mention them more than once.

and no-go theorem [270, 272–274, 297]. Aaronson [272] argued in a blog post that the reasoning statements made by the friends are invalid at the end of the protocol as the brains of the friends have been 'Hadamarded' in the process. It is not clear to us whether this argument claims that (i) only reasoning or (ii) also mere experiencing of the measurement result by the unitarily-modeled quantum observers is invalid. Case (i) has been criticized by del Rio and Renner [298], and anyway does not apply to the GHZ–FR paradox, as only 'classical' agents reason. Case (ii) could correspond to rejecting Universality of Quantum Theory, but may also suggest a relativity of observed events, depending on its exact interpretation. Healey [270] argues that an additional assumption of Intervention Sensitivity is required for reasoning about outcomes obtained by quantum agents, and Drezet [273] asserts that in a Bohrian interpretation reasoning by quantum agents may be problematic. Vilasini and Woods [24] suggest that the arguments of Refs. [270, 297] also refer to the post-selection step, but this depends on the exact reading of these two references. Criticisms of post-selection and reasoning by quantum agents do not apply to the GHZ–FR paradox as no post-selection or reasoning by quantum agents is needed. Araújo [297] seems to argue that both collapse and no-collapse modeling of a measurement is needed for the FR argument. However, we note that the state-update rule is not needed [274], and only unitary modeling and the Born rule for predictions are used. Namely, each use of the Born rule in the FR involves only measurements that can be performed jointly.[20] Fortin and Lombardi [274] argue that the FR argument is either illegitimate or uninteresting and simply another proof of contextuality as agents make reasoning statements that correspond to different contexts when mapping the FR protocol to a Bell nonlocality scenario, but do not respond to how agents can then use quantum theory consistently in the presence of superobservers.

Secondly, the FR paradox has been analysed from the viewpoint of specific interpretations of quantum theory [264, 273, 283–285, 288, 299–302], and arguments have been put forward for which assumption of the FR paradox to refute. The FR paradox has been studied in Bohmian and Bell–Bohmian theories in Refs. [273, 283, 284], in a relational quantum theory setting in Ref. [299], from a QBist perspective in Ref. [288], and in the consistent-histories interpretation in Ref. [302]. An overview of the FR paradox in different interpretations of quantum mechanics can be found in Ref. [264].

- The (Bell–)Bohmian analyses of the FR paradox [273, 283–285] argue that the agents apply the Born rule (Q) wrongly, as they do not take into account the full

[20] We thank David Schmid and Yìlè Yīng for discussions on this matter.

state, i.e. the pilot wavefunction. But still, one would like the agents to be able to make correct predictions, even if they cannot take an outsider's view. We have provided a possible way to phrase their predictions correctly in Section 3.2.2 under 'Reject Born Compatibility'.

- The relational and QBist perspectives propose to refute FR's consistency assumption (C), allowing for different agents to assign different truth values to statements [264, 288, 299]. One may argue that also the consistent-histories approach to the FR [302] refutes (C) [264], as inferences are made between inconsistent sets of histories. When refuting (C), one would still like to know when one agent can trust another agent's statement or reasoning. For example, in the GHZ–FR paradox, one can imagine that Ursula's, Valentina's, and Wigner's measurements are all performed by the same agent (sequentially in their proper time), who then makes statements similar to those of Ursula, Valentina, and Wigner: must this agent then not trust their own statement? In the QBist approach to FR by DeBrota, Fuchs, and Schack [288], it is for example argued that when to trust another agent's statement can be decided on a case-by-case basis. Our GHZ–FR no-go theorem shows that even classical communication cannot be trusted when allowing different truth values for different agents, and we have argued that perhaps it is easier to instead restrict the validity of statements using the Born rule for predictions, while holding on to consistency between agents. The fact that no quantum agents need to reason in the GHZ–FR paradox is crucial for this argument, and poses sharper restrictions on building consistent reasoning principles.[21]
- Other works [32, 301, 303–305] argue for refuting the existence of superobservers because of decoherence or absolute collapse, rendering the FR protocol – involving superobservers – impossible, even in principle. However, decoherence may be part of an explanation of the quantum-to-classical transition, but it is not a full answer to the measurement problem [6, 306, 307]. Absolute collapse requires an explanation of how or why the collapse happens, as in spontaneous collapse theories [308, 309] or gravity-induced collapse [310, 311]. Experiments may shed light on whether collapse theories, which have different dynamics than ordinary quantum theory, are likely to be correct [312–316].

[21] Whether the relative-facts approach of Ref. [299], which refutes (C) and argues to not believe the reasoning of the quantum agents, resolves the paradoxes presented in this chapter or not depends on its exact interpretation. For a more elaborate discussion see Appendix E of Ref. [19].

Thirdly, there is some work towards responding to Renner's challenge of building a consistent set of reasoning principles for quantum theory that evades contradictions like the FR paradox [24, 317–319]. Here we note that Vilasini and Woods [24] propose that one be explicit about which Heisenberg cuts are used in each statement concerning outcomes of performed measurements. This leaves open the question of the physical meaning of Heisenberg cuts, and why statements should be formulated as dependent on them? Is it merely a choice of what to model inside and outside the quantum theory? If so, then still desirably no contradictions about the outcome values as in the (GHZ–)FR paradox should arise when making different choices, if it is merely a choice. However, a resolution à la Relativity of Observed Events proposes a reason of why such mentioning would be necessary: it refers to the statement being conditional on physically obtaining (a copy of) the outcome record(s) one is talking about; a prediction by agent A is meant for when A actually obtains those outcome records (and the protocol be performed truthfully per agent A's knowledge). Moreover, we also note that the GHZ–FR no-go theorem invalidates the set of reasoning principles from Ref. [319]. For a proof see Appendix E.2 of Ref. [19].

Our work belongs in the first category, as it provides a stronger no-go theorem than the FR. It also touches on the third category by suggesting ways to resolve paradoxes involving superobservers in quantum theory.

3.4.2 Related work on the GHZ–FR paradox

A protocol similar to the one underlying the GHZ–FR paradox has been proposed by Żukowski and Markiewicz [32] and by Leegwater [26]. While the measurements performed in the protocols (without reasoning steps) in Refs. [26, 32] are the same, the exact way in which a contradiction is obtained differs, as do the ensuing analyses and no-go theorems. In Ref. [32], an explicit motivation for how the different measurement contexts that lead to a contradiction are chosen and considered together is not provided. In Ref. [26] this motivation comes from simultaneity in special relativity. However, in cases where the motivation from special relativity of Ref. [26] cannot be applied, the GHZ–FR paradox still yields a contradiction; for more detail see Appendix F of Ref. [19].

Going beyond Ref. [26], we investigate the relationship of the GHZ–FR protocol to the FR paradox more thoroughly, and expose the hierarchy of their underlying nonlocality models. Unlike Refs. [26, 32] our agreement no-go theorem requires no absolute outcomes, due to the reasoning part of our protocol being different. This important

weakening allowed us to propose the Relativity of Observed Events resolution as replacing Absoluteness of Observed Events in Sections 3.2.2 and 3.2.3. Unlike these works [26, 32], we move away from having a bird's-eye view on the protocol as an outsider, and instead consider the viewpoints from observers inside the protocol who try to make predictions and reason, as was the goal in the original FR paradox. To relate to the notion of *Born Inaccessible correlations*[22] of Ref. [27], we acknowledge that there will always be correlations inaccessible to any agent. However, we assign the label 'inaccessible' as we can take the outsider's bird's-eye view on the protocol; it might be impossible for agents inside to have this knowledge in all situations. Were we to make a no-go theorem using the notion of Born Inaccessible correlations, we would find a no-go theorem similar to the GHZ–FR truth no-go theorem, replacing Born Compatibility with Born Inaccessible correlations. For a further discussion of the works [26, 32] see Appendix F in Ref. [19].

A similar relational resolution as Relativity of Observed Events in the context of Relational Quantum Mechanics [320, 321] was discussed by Cavalcanti, Di Biagio and Rovelli [300] for a scenario involving the same measurement protocol as the GHZ–FR paradox, based on Refs. [32, 322]. Our resolution per Relativity of Observed Events accords with their proposed resolution, except that our proposal does not invoke the existence of cross-perspective links [323]. Furthermore, as noted by Cavalcanti and Wiseman in Ref. [13], to resolve the Local Friendliness no-go theorem, by giving up on AOE for a notion of relative events one can maintain compatibility with Leibniz's principle.

3.4.3 Relation to the Local Friendliness no-go theorem

The Local Friendliness no-go theorem [12–14, 27] is based on a protocol similar to the FR's but is phrased in a device-independent fashion. It uses the Born rule only for outcomes that are obtained by us (i.e. that are available after the protocol), but in addition to AOE it also requires a notion of free choices satisfying 'Local Agency', a locality constraint not invoking hidden variables as in Bell's local causality. If in the GHZ–FR paradox of Section 3.1 we give the superobservers a choice between performing their measurement or asking their respective observers for their outcomes, we obtain a Local Friendliness scenario [12, 14]. In Chapter 4 we clarify how also contextuality scenarios may be used to produce a Local Friendliness no-go theorem.

[22] Born Inaccessible Correlations is defined in Ref. [27] as: "It requires that when two measurements are made in parallel, the two outcomes always arise with a joint frequency given by the Born rule, even if no single observer could access both outcomes, even in principle."

3.5 Conclusion

We conclude by summing up the main results regarding the GHZ–FR paradox and outline some future research directions.

3.5.1 Summary

Motivated by the question of what the FR paradox [11] implies for quantum foundations, we presented a similar but strictly stronger paradox, the GHZ–FR paradox, based on the strongly contextual GHZ–Mermin model [276, 277]. Outlining carefully the required assumptions, we obtained two no-go theorems of increasing strength, which we named the GHZ–FR truth and agreement no-go theorems. The easiest to grasp is the GHZ–FR truth no-go theorem, whose assumptions are summarized in Table 3.2. The agreement no-go theorem weakens Absoluteness of Observed Events replacing it by Personal Knowledge and Classical Agreement; see Table 3.3: instead of *absolute* outcome assignments to performed measurements, it posits outcome assignments that may a priori differ per agent, subject to the condition that classically communicating agents must agree on their outcome assignments.

GHZ–FR truth no-go theorem	
Absoluteness of Observed Events (AOE)	Every performed measurement has an absolute, single-valued outcome.
Born Compatibility	Outcomes assigned (per AOE) to a set of measurements that an agent A believes she can gather jointly in a memory must be compatible with A's use of the possibilistic Born rule.
Universality of Quantum Theory	Superobservers exist (in principle) and quantum theory is correct.

Table 3.2: Assumptions of the GHZ–FR truth no-go theorem; more precisely stated assumptions can be found in Section 3.2.1.

Pursuing the route of rejecting AOE or Born Compatibility, as discussed in Section 3.2.2, the agreement no-go theorem suggests a principle that we call Born Practicality, defined in Section 3.2.3, meaning that an agent A can use the Born rule only to formulate predictions that condition on A actually learning about all involved outcomes. This principle can be seen as an agent-based extension of Peres's dictum [39] 'Unperformed experiments have no results':

> 'Unperformed experiments have no results, and unknown results have no values'.

GHZ–FR truth no-go theorem	GHZ–FR agreement no-go theorem	
Absoluteness of Observed Events	Personal Knowledge	Every performed measurement can be assigned a single-valued outcome by an agent who believes they are able to gather that outcome.
	Classical Agreement	The outcome assignments of classical agents who communicate classically must agree on their overlap.

Table 3.3: The GHZ–FR agreement no-go theorem is based on the same assumptions as the GHZ–FR truth no-go theorem, but with Personal Knowledge and Classical Agreement replacing Absoluteness of Observed Events.

One may view this as a natural quantum-vs-classical statement. Therefore, as a possible resolution to extended Wigner's friend paradoxes we propose Born Practicality as a principle governing agents' use of quantum theory in a world where Universality of Quantum Theory holds. In a relational framework, rejecting AOE, Born Practicality can be implemented by having single-valued outcomes only defined relative to the experimenter and other agents who learn about the outcome, which we named Relativity of Observed Events in Section 3.2.2. In a hidden-variable framework, rejecting Born Compatibility, one may instead have hidden variables determining absolute, single-valued outcomes of all performed measurements but agents must follow Born Practicality for correct predictions of the Born rule. Of course, one may also question the existence of superobservers, or simply allow for classically communicating agents to have inconsistent truth assignments. The latter, however, seems impractical, as we argued in Sections 3.2.2 and 3.4.

3.5.2 Outlook

To conclude, our work opens up various directions for future work. In upcoming work, we will discuss the implications of the GHZ–FR paradox for the measurement problem, for different interpretations of quantum theory, and how it suggests a resolution to Renner's challenge of providing a consistent set of reasoning rules in quantum settings. Another intriguing question is what implications the GHZ–FR paradox and other extended Wigner's friend paradoxes have for the discussion on the epistemic or ontic nature of quantum states [38, 231, 259, 324] or ontologies of quantum mechanics. A deeper investigation of the relationship between Relativity of Observed Events or Born Practicality and other proposed resolutions of extended Wigner's friend paradoxes, including bubbles and quantum influences [325, 326], the relational approach of 'fact-nets' [327], and quantum reference frames [328, 329] is left for future work.

Another interesting avenue to explore is to combine Wigner's friend with other nonclassical quantum features such as KS noncontextual models that are not nonlocality models, steering [330, 331], incompatibility [332], generalized contextuality [43, 44], coherence [333, 334], etc. We do so for Kochen–Specker, generalized noncontextuality and the Pusey–Barrett–Rudolph argument [38] in the next chapters.

— 4 —

Extended Wigner's friend arguments and noncontextuality

As mentioned in Section 2.3, previous extended Wigner's friend (EWF) arguments are primarily based on Bell nonlocality. In the next four sections, we present links between extended Wigner's friend arguments and other notions of nonclassicality, such as Kochen–Specker (KS) and generalized noncontextuality, and introduce a class of new EWF arguments.

Specifically, in Section 4.1, we present an extended Wigner's friend argument using a single system, based on KS noncontextuality, by suitably generalizing an assumption used in previous EWF arguments. Next, in Section 4.2, we focus on the Local Friendliness (LF) no-go theorem. We show how constraints on scenarios, arising from the Local Friendliness assumptions, are mathematically related to Kochen–Specker noncontextuality requirements on the relevant measurements. This allows to find many more inequalities that the LF polytope (i.e. the operational correlations fulfilling the Local Friendliness assumptions) satisfies. Moreover, in Section 4.3 we construct scenarios violating Local Friendliness based solely on Kochen–Specker contextuality.

These findings raise the question of whether a yet more general notion of nonclassicality, generalized contextuality, can be used to derive EWF arguments. In Section 4.4, we answer this question affirmatively, under the condition that one generalizes the Local Friendliness assumptions to a more spacetime-independent form. Namely, the LF assumptions always refer to a background spacetime where one assumes the causality of Special or General Relativity to hold. However, one might be interested in an EWF no-go theorem that does not make use of such background, for example when working with background-independent quantum gravity theories, or to generalize to cases where the precise locations

are unknown. By doing so, we uncover that behind the intuition of Local Friendliness and EWF arguments more generally is a principle of no fine-tuning, a well-known concept in quantum foundations, under the assumption that the Absoluteness of Observed Events holds.

More specifically, the Local Friendliness assumptions consist of two main parts: Absoluteness of Observed Events and Local Agency. Local Agency captures a notion of locality, that each choice can only be correlated with events in its future. Local Agency can be motivated from the postulate that no physical system can (locally) travel faster than the speed of light in special and general relativity. However, it can also be motivated from the empirical finding of no-signaling. No-signaling here means that no physical message (a choice x) can be sent faster than the speed of light; all events $a, b, ...$ outside of x's future lightcone must be uncorrelated with x. Local Agency can be motivated by requiring that no-signaling still holds for those joint correlations existing by Absoluteness of Observed Events, even when it cannot be verified by a single agent. Our generalization, named Operational Agency, captures the idea that operational equivalences, i.e. empirically observed independencies of choices, must still hold for the joint correlations that exist by Absoluteness of Observed Events even when the relevant events are not brought together (i.e. not observed by a single agent). In other words, it expresses a notion of no-conspiracy, or no fine-tuning: observed independencies must remain true even when unobserved (but whose correlations exist when assuming Absoluteness of Observed Events). Quantum theory violates Local Friendliness and Operational Friendliness, assuming its universality in that superobservers exist.

Beyond simply providing additional proofs of Local Friendliness, these works contribute by identifying the broader nonclassical structure underlying such (LF) results and distinguishing essential assumptions from features specific to the original LF construction. Moreover, they situate the LF assumptions within established notions of noncontextual classical explanation, clarify their dependence on background spacetime structure and extend EWF arguments to scenarios outside the standard multipartite Bell framework.

Rather than providing a single discussion at the end, we will provide brief discussions in the different sections of this chapter on the relevant notions introduced, when necessary.

As we are concerned with the Local Friendliness no-go theorem in this chapter and the next, we follow the conventions in literature and introduce a notational difference between empirical probability distributions that can be obtained by a single agent from

the collected data, denoted by $\wp$, and theoretical probability distributions required to exist given certain assumptions, but potentially not observable by any agent, denoted by p.

Section 4.1 is based on Ref. [20], Sections 4.2 and 4.3 on Ref. [21] and Section 4.4 on Ref. [22].

4.1 An extended Wigner's friend paradox based on KS contextuality

Thus far, we have seen how EWF arguments are based on Bell nonlocality models. As KS contextuality can be seen as a generalisation of Bell nonlocality, one can ask whether one can construct an EWF argument that involves measurements only on a single system based on KS contextuality. In this subsection we answer this question affirmatively, and produce an EWF argument based on the 5-cycle KCBS contextual model [237].

4.1.1 Assumptions

A number of assumptions we consider here are common to most EWF arguments, the first two being the Universality of Quantum Theory and Absoluteness of Observed Events [12–14], introduced in Chapters 2 and 3, which we repeat here.

Assumption 3 (Universality of Quantum Theory)**.** A sealed lab in which an observer performs a measurement on some system can be described as a closed quantum system that evolves unitarily, upon which a superobserver can apply arbitrary quantum operations, including measurements.

Assumption 4 (Absoluteness of Observed Events (AOE))**.** An observed event is an absolute single event, not relative to anything or anyone.

Another standard assumption for EWF no-go theorems is that the outcome for any single measurement (or outcomes for measurements done in parallel, if those outcomes are jointly observable) occurs at frequencies obeying the Born rule, in accordance with operational quantum theory. Here we only need the *possibilistic Born rule*, demanding that such outcomes never occur if the Born rule assigns probability 0 to them, and that otherwise, they sometimes occur. Often this assumption is not stated explicitly, as it simply concerns the experimental validity of quantum theory.

We will make use of all the assumptions introduced above. Our no-go theorem also makes use of one novel assumption called *Commutation Irrelevance*, which will be introduced in due time.

4.1.2 The 5-cycle noncontextuality scenario

We briefly recap the 5-cycle noncontextuality no-go theorem from Ref. [237, 335, 336], as presented in Section 2.2.1, and provide a quantum realization upon which our EWF argument will be based. Recall that a Kochen–Specker noncontextual assignment [41] for a set of measurements is a deterministic assignment of outcomes to those measurements, such that the assignment for each measurement is independent of which other compatible measurements are performed jointly with it.

Consider five binary-outcome measurements $\{M_i\}_{i\in\{1,2,\ldots,5\}}$, where the pairs $\{(1,2)$, $(2,3)$, $(3,4)$, $(4,5)$, $(5,1)\}$ are jointly measurable, with compatibility graph of Fig. 4.1 and outcome probabilities satisfying

$$\wp(1,1|1,2)=0, \tag{4.1a}$$
$$\wp(0,0|2,3)=0, \tag{4.1b}$$
$$\wp(1,1|3,4)=0, \tag{4.1c}$$
$$\wp(0,0|4,5)=0, \tag{4.1d}$$

and

$$\wp(0,1|5,1)\neq 0. \tag{4.2}$$

A deterministic KS assignment of the outcome of measurement M_i as m_i, Eq. (4.1) satisfies

$$m_1=1\Rightarrow m_2=0\Rightarrow m_3=1\Rightarrow m_4=0\Rightarrow m_5=1. \tag{4.3}$$

However, Eq. (4.2) tells us that in some runs of the protocol, the outcome of M_5 is 0 and the outcome of M_1 is 1. Thus, there is no Kochen–Specker noncontextual assignment for these runs.

The set of correlations in Eq. (4.1) and Eq. (4.2) has a quantum realization as follows. Consider a qutrit prepared in the state

$$|\eta\rangle:=\sqrt{\frac{1}{3}}(|0\rangle+|1\rangle+|2\rangle)\equiv\sqrt{\frac{1}{3}}(1,1,1)^T, \tag{4.4}$$

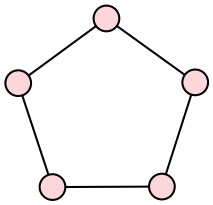

Figure 4.1: The compatibility graph of the 5-cycle scenario. The pairs $\{(1,2),(2,3),(3,4),(4,5),(5,1)\}$ are jointly measurable.

where T denotes transposition, and define measurements

$$M_i := \{|v_i\rangle\langle v_i|, \mathbb{1} - |v_i\rangle\langle v_i|\}, \tag{4.5}$$

with the states $|v_i\rangle$ defined as

$$\begin{aligned} |v_1\rangle &= \sqrt{\frac{1}{3}}(1,-1,1)^T, \\ |v_2\rangle &= \sqrt{\frac{1}{2}}(1,1,0)^T, \\ |v_3\rangle &= (0,0,1)^T, \\ |v_4\rangle &= (1,0,0)^T, \\ |v_5\rangle &= \sqrt{\frac{1}{2}}(0,1,1)^T. \end{aligned} \tag{4.6}$$

We take outcome 0 of a given measurement to correspond to $\mathbb{1}-|v_i\rangle\langle v_i|$ and outcome 1 to correspond to $|v_i\rangle\langle v_i|$. In this case, Eqs. (4.1a)-(4.1d) are satisfied (regardless of the state that the measurements act on) while $\wp(0,1|5,1) = \left|\langle\eta|(\mathbb{1} - |v_5\rangle\langle v_5|)|v_1\rangle\right|^2 = 1/9 \neq 0$, leading to Eq. (4.2). Hence, quantum theory is not consistent with noncontextuality.

4.1.3 An EWF scenario with measurements on a single system

We now leverage these 5-cycle correlations to construct an EWF no-go theorem. The argument involves five friends (denoted F_1, ..., F_5) who perform the five measurements M_i defined in Eq. (4.5) sequentially on a single system S prepared in the state $|\eta\rangle$ of Eq. (4.4), and a superobserver Wigner who undoes the first three of these measurements at particular times during the protocol.

By the assumption of Universality of Quantum Theory, each observer can be modeled as a quantum system—and indeed, as a qubit, since only two orthogonal (coarse-grained) states of the observer are relevant to the argument. If each measurement M_i is performed

in a sufficiently isolated environment, Universality of Quantum Theory further implies that it can be described by a unitary. In particular, if the measurements are performed in a manner that has minimal disturbance on system S, then each measurement M_i can be modeled by a CNOT gate U_{M_i} between the system S and the agent A_i, namely,

$$\begin{aligned} U_{M_i} &:= \mathrm{C}NOT_{SF_i} \\ &= (\mathbb{1} - |v_i\rangle\langle v_i|)_{\mathrm{S}} \otimes \mathbb{1}_{\mathrm{F}_i} + |v_i\rangle\langle v_i|_{\mathrm{S}} \otimes (|1\rangle\langle 0| + |0\rangle\langle 1|)_{\mathrm{F}_i}. \end{aligned} \tag{4.7}$$

where $|0\rangle_{F_i}$ is the coarse-grained state of F_i having observed outcome 0 and $|1\rangle_{F_i}$ is for having observed outcome 1. We follow a standard abuse of notation by letting $|0\rangle_{F_i}$ also denote the coarse-grained 'ready' state of the observer—that is, the initial state of each observer (prior to the measurement).

The superobserver Wigner is assumed to have perfect quantum control over the joint system of F_1, F_2, F_3, and S. In particular, we assume he has the (extreme) technological capabilities to implement the inverse of the unitary operations for the first three friends' measurements.

The exact sequence of operations is shown in the unitary circuit representation of the protocol in Fig. 4.2. Specifically, M_1 is the first measurement on S, followed by M_2, after which the superobserver undoes the first measurement by applying the unitary $U_{M_1}^\dagger$. Then M_3 is performed, followed by the undoing $U_{M_2}^\dagger$, followed by M_4, followed by the undoing $U_{M_3}^\dagger$, followed by M_5.

From Absoluteness of Observed Events, each agent A_i observes an absolute outcome, denoted f_i, during their measurement M_i. We label the $|v_i\rangle\langle v_i|$ outcome by $f_i = 1$ and the other by $f_i = 0$.

4.1.4 No-go theorem

Measurements M_1 and M_2 are done immediately in sequence without any processes occurring in between, so the joint outcome (f_1, f_2) can be directly observed (up until the time when $U_{M_2}^\dagger$ is performed).[1] The Born rule predicts that

$$\wp(f_1 = 1, f_2 = 1) = 0, \tag{4.8}$$

[1] In fact, even though the records of the outcomes f_1 and f_2 may be erased later, it is nevertheless possible to obtain empirical evidence that the joint outcome (f_1, f_2) agrees with the Born rule predictions. For example, we could add an extra instruction to friend F_2 in the experimental protocol: if she observes $f_1 = f_2 = 1$, in contradiction with Eq. (4.8), she should open the door of her lab and request that the experiment be halted immediately. Furthermore, in the case where (f_1, f_2) is as expected, F_2 could write a note stating "agrees with the Born rule" and pass it to Wigner through a tiny slit in her lab door that is small enough and opened for a short enough time so that no other interaction between the inside

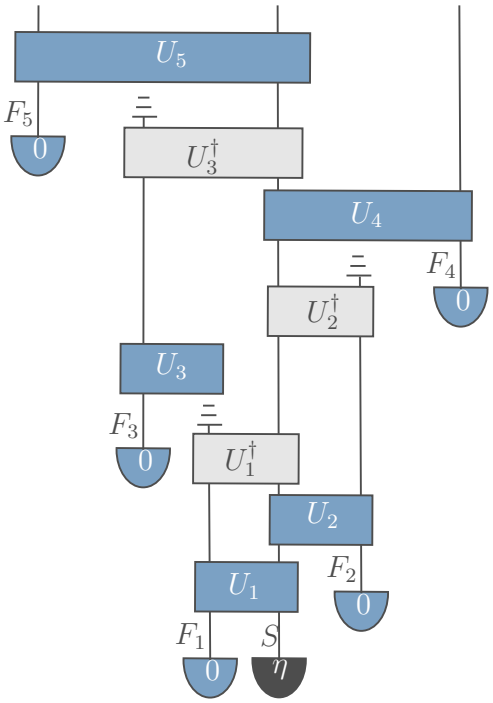

Figure 4.2: **Schematic depiction of the scenario of Section 4.1.3**. Five friends $F_1, \ldots, F_5$ perform the five measurements from a 5-cycle noncontextuality no-go theorem on a system S, while a superobserver Wigner undoes the first three of these measurements at particular times in between. Each measurement M_i is modeled unitarily as U_{M_i}, following the assumption of Universality of Quantum Theory. The unitaries from U_{M_2} to $U_{M_3}^\dagger$ are collectively denoted as U. In each measurement M_i, agent F_i obtains an absolute outcome f_i, following the assumption of Absoluteness of Observed Events. Here the three small horizontal lines $\equiv$ on a vertical line (a wire) means the system corresponding to the wire can be discarded, i.e. no more operations must be performed on it.

just as in Eq. (4.1a), the correlation for the $(1, 2)$ pair of measurements in the 5-cycle noncontextuality scenario. Moreover, this Born rule prediction—and those that follow—can be made without appealing to the projection postulate, as we explain later. This correlation f_1, f_2 is available to observers up until $U_{M_1}^\dagger$ is applied (which happens after measurement M_2).

Measurements M_2 and M_3 are not done immediately in sequence; rather, the unitary $U_{M_1}^\dagger$ is applied in between them. Nevertheless, the joint outcome (f_2, f_3) is still directly

and outside of the lab can occur. This note can be retained by Wigner even if the records of (f_1, f_2) are later erased, and the rest of the experimental protocol can proceed unaffected. The inclusion of this note is analogous to the one in Deutsch's version of Wigner's friend [263, Sec. 8], the one in Ref. [337, Sec. IV.B], and the one in Ref. [33, Sec. 4.2.2].

observable (up until $U^\dagger_{M_2}$ is performed). The Born rule predicts that

$$\wp(f_2 = 0, f_3 = 0) = 0, \tag{4.9}$$

just as in Eq. (4.1b), the correlation for the $(2, 3)$ pair of measurements in the 5-cycle noncontextuality scenario. This is the case because the unitary $U^\dagger_{M_1}$ commutes with the M_3 measurement (as shown explicitly in Appendix E of Ref. [20]), and so it does not affect the Born rule prediction for $\wp(f_2, f_3)$.

Similarly, the Born rule predicts that

$$\wp(f_3 = 1, f_4 = 1) = 0, \tag{4.10}$$

$$\wp(f_4 = 0, f_5 = 0) = 0, \tag{4.11}$$

where both refer to joint outcomes that can be directly observed, and the only process happening between each pair of measurements is a unitary that commutes with those measurements (and so cannot affect their statistics). Eqs. (4.10) and (4.11) are analogous to Eqs. (4.1c) and (4.1d), respectively, in the 5-cycle noncontextuality scenario.

However, the same argument cannot be applied to the joint outcome (f_1, f_5), since it necessarily cannot be observed by any observer, even in principle. This is because the outcome f_1 is erased by the unitary $U^\dagger_{M_1}$ (meaning all outcome records of f_1, direct and indirect, are physically erased) prior to the outcome f_5 being generated. However, if the process $U := U^\dagger_{M_3} U_{M_4} U^\dagger_{M_2} U_{M_3} U^\dagger_{M_1} U_{M_2}$ would be applied after M_5 instead of before M_5, the correlation (f_1, f_5) would be available to a single agent. Consequently, even though the process U, which is done between M_1 and M_5, commutes with M_5, the assertion that it does not affect the correlation between M_1 and M_5 is *not* guaranteed by the correctness of operational quantum theory (in particular, it does not follow from our assumption of Possibilistic Born rule). Different interpretations of quantum theory may make different predictions for unobservable correlations, while still being consistent with operational quantum theory, as was highlighted already in, e.g., Ref. [27].

Therefore, to complete the argument, we require an assumption that the commutation relation $U_{M_5} U U_{M_1} = U U_{M_5} U_{M_1}$ implies $p_{U_{M_5} U U_{M_1}}(f_1, f_5) = p_{U U_{M_5} U_{M_1}}(f_1, f_5)$, where the former denotes the joint distribution when U is performed between M_1 and M_5, and the latter denotes the joint distribution when U is performed after M_5. Let us call this assumption *Commutation Irrelevance*.[2]

[2]We note that someone with operationalist leanings might object to this assumption on the grounds that it constrains correlations that are unobservable in principle. However, we note that it does not

Assumption 5 (Commutation Irrelevance)**.** Any unitary process performed between two measurements does not affect the correlations between their outcomes, provided that the unitary commutes with at least one of the two measurements—*even* if the outcomes of the two measurements are not jointly observable to any observer.

In the latter case, where U is carried out *after* M_5, the correlation $\wp(f_1, f_5)$ is observable, and so is constrained by operational quantum theory to be $p_{UU_{M_5}U_{M_1}}(f_1 = 1, f_5 = 0) \neq 0$. Consequently, assuming Commutation Irrelevance, the unobservable correlations for the actual experiment (where U is done between M_1 and M_5) are constrained in the same way:

$$p(f_1 = 1, f_5 = 0) \neq 0. \tag{4.12}$$

This is analogous to Eq. (4.2) in the 5-cycle noncontextuality scenario.

Eqs. (4.8) to (4.11) imply that the actually observed outcomes must satisfy

$$f_1 = 1 \Rightarrow f_2 = 0 \Rightarrow f_3 = 1 \Rightarrow f_4 = 0 \Rightarrow f_5 = 1. \tag{4.13}$$

However, Eq. (4.12) tells us that the event $(f_1 = 1, f_5 = 0)$ occurs in some runs of the protocol.

Thus, we have a contradiction and consequently a no-go theorem against the conjunction of our assumptions: Universality of Quantum Theory, Absoluteness of Observed Events, possibilistic Born Rule and Commutation Irrelevance.

In Appendix D of Ref. [20], this scenario is extended to an infinite family of EWF paradoxes based on n-cycle proofs of the failure of Kochen–Specker noncontextuality for $n \geq 4$. There we also introduce a unified description of possibilistic n-cycle contextuality models for both even and odd n.

4.1.5 Discussion and outlook

How well motivated is Commutation Irrelevance?

No-go results are only as interesting as their underlying assumptions are compelling. Hence, it is necessary to question whether the assumption of Commutation Irrelevance is well-motivated. This is the key assumption beyond the standard assumptions common to EWF arguments, and plays a role analogous to the auxiliary assumptions of Timing

suffice to retreat to the position that 'unperformed experiments have no results' [39], since the relevant correlations in an EWF argument refer to outcomes of *actually performed measurements*. This point was already made in e.g., Refs. [12, 13, 27].

Irrelevance or Local Agency in the Frauchiger–Renner and Local Friendliness arguments, respectively.

In fact, the assumption of Commutation Irrelevance is a minor extension of the assumption of Timing Irrelevance [27] required in the Frauchiger–Renner (and the Pusey–Masanes [338, 339]) arguments. Timing Irrelevance is the assumption that the correlations obtained in two quantum circuits whose only difference is the timing of the measurements must be identical (even if the difference in timings matters for the question of whether or not the correlations in question are in principle observable). Timing Irrelevance is a special case of Commutation Irrelevance, since the identity channel (representing the action of waiting) commutes with any operation.

Both Timing Irrelevance and Commutation Irrelevance can be motivated by the assumption that quantum theory is *complete* in the sense that there is no deeper theory or deeper set of facts (like hidden variables) that determine what measurement outcomes occur. This assumption is endorsed by researchers sympathetic to the Copenhagen school of thought. On the other hand, as argued in Ref. [27], for interpretations that violate Completeness, or researchers who are less certain that quantum theory is the deepest possible description of nature, Timing Irrelevance and Commutation Irrelevance may not be compelling assumptions.

One explicit example of an interpretation that violates Commutation Irrelevance is Bohmian mechanics. This follows from the fact that Bohmian mechanics violates even the weaker assumption of Timing Irrelevance, as discussed in Ref. [27]. One can also see Bohmian mechanics's violation of Commutation Irrelevance in the scenario discussed here by noticing that Bohmian mechanics reproduces quantum predictions for *observable* correlations. For our scenario (where U is done between M_1 and M_5), Bohmian mechanics reproduces the observable predictions that $f_1 = 1 \Rightarrow f_2 = 0 \Rightarrow f_3 = 1 \Rightarrow f_4 = 0 \Rightarrow f_5 = 1$. Hence, $p_{U_{M_5}UU_{M_1}}(f_1 = 1, f_5 = 0) = 0$ holds. However, for the case where U is done after M_5, so that the outcomes of M_1 and M_5 can be jointly observed, Bohmian mechanics reproduces the quantum prediction for this correlation, namely, that $p_{UU_{M_5}U_{M_1}}(f_1 = 1, f_5 = 0) \neq 0$. So in Bohmian mechanics, $p_{U_{M_5}UU_{M_1}}(f_1 = 1, f_5 = 0) \neq p_{UU_{M_5}U_{M_1}}(f_1 = 1, f_5 = 0)$.

Related works There are a few related works that construct EWF arguments on a single system [15, 18, 340]. Refs. [18, 340] implement constructions similar to ours based on contextual correlations, but do not identify or motivate the critical assumption that (we will argue) is required. In addition, Ref. [340] applies the projection postulate

to processes that are *also* treated unitarily in the argument, which (as mentioned just above) undermines the apparent paradox.

Ref. [15] also builds an EWF argument on a single system—one that is not based on contextual correlations, but rather on an assumption of a kind of linearity [15]. This assumption can be criticized on interpretation-independent grounds [27]. Although the novel assumption required in our construction can also be criticized and would be rejected in some interpretations (as we discuss later), we argue that it may be motivated within others (and in particular, the same ones under which Frauchiger–Renner's assumptions are motivated). Still, understanding the problems with the linearity assumption in Ref. [15] gives interesting insights—e.g., into the nature of quotiented representations of quantum processes [27], and so is worth study. Similarly, we hope that our no-go theorem will help provide insights into the metaphysical assumptions going into the Frauchiger–Renner argument and related arguments (with consequent implications for interpretations of quantum theory).

4.2 Deriving Local Friendliness inequalities from KSNC constraints

In this section, we show how techniques from the study of Kochen–Specker noncontextuality can be used to derive many Local Friendliness inequalities in a wide range of Local Friendliness scenarios, including multipartite systems and sequential measurements. We do so by considering Kochen–Specker inequalities relative to *subgraphs* of the compatibility graph naturally associated to the Local Friendliness scenario in question.

4.2.1 Technical consequences of the Local Friendliness assumptions for LF scenarios

We first introduce some terminology regarding the N_A, N_B-setting LF scenario of Section 2.5. Recall that in this scenario, we have two friends, Charlie and Debbie, performing a measurement, which we denote by A_1 and B_1, after which Alice and Bob each choose $x \in \{1, 2, \ldots, N_A\}, y \in \{1, 2, \ldots, N_B\}$. For $x = 1$, Alice opens Charlie's lab and reveals his outcome a_1 to the world. For $x \in \{2, \ldots, N_A\}$, Alice performs some operation on Charlie's lab (for example undoing Charlie's measurement in the quantum scenario), after which Alice instructs Charlie to perform another measurement on his system; which

measurement she instructs him to perform is determined by $x \in \{2, \ldots, N_A\}$. Charlie's new measurement is denoted A_x, and his respective outcome is denoted a_x. For instance, if $x = 2$, the new measurement performed by Charlie is A_2 with outcome a_2. After Charlie has performed his new measurement, Alice reveals his outcome a_x to the world. Bob and Debbie act analogously.

At the end of the experiment, we collect statistics for a_x, b_y, x, and y, which constitute the empirical probabilities $\wp(a_x, b_y|x, y)$. These probabilities constitute the empirical correlations (or behaviors) in this LF scenario. For simplicity, we also denote these empirical probabilities as $\wp(a_x, b_y)$ without the explicit conditionals since the conditionals are actually redundant — the empirical correlation between a_x and b_y is only defined in the runs with the corresponding choices x and y.

We now look in some more detail at the technical consequences of the Local Friendliness assumptions for the (N_A, N_B)-setting LF scenario.

In each individual run of the LF scenario, either two, three, or four measurements are performed, depending on Alice's and Bob's choices. For example, when $x = 1, y = 1$, only two measurements are performed while when $x = 2, y = 2$, four measurements are performed in a single run. Thus, depending on the choices, in each run, either two, three, or four outcomes are observed by the agents. AOE implies that for a given value of x and a given value of y, there exists a joint distribution over the outcomes observed in the corresponding runs. For example, for $x = i, y = 1$, measurements A_1, A_i, B_1 are performed and their respective outcomes exist in a single run. By AOE, for these runs, one can assign a well-defined joint distribution over their outcomes a_1, a_i, b_1, which we denote as $p(a_1, a_i, b_1|x{=}i, y{=}1)$.[3] By the experimental setup, the empirical correlation $\wp(a_i, b_1)$ is its marginal, i.e., $p(a_1, a_i, b_1|x{=}i, y{=}1) = \wp(a_i, b_1)$. For all given pairs of values x, y of choices, we have the following respective joint distributions:

$$p(a_1, b_1|x = 1, y = 1), \tag{4.14a}$$

$$p(a_1, a_i, b_1|x = i, y = 1), \tag{4.14b}$$

$$p(a_1, b_1, b_j|x = 1, y = j), \tag{4.14c}$$

$$p(a_1, a_i, b_1, b_j|x = i, y = j), \tag{4.14d}$$

[3] Here we follow the common abuse of notation in the literature that the same letter such as a_1 is used to denote both the random variable representing the outcome and the value of the random variable. If one wishes to write the distribution in a more explicit way, one could use $\mathcal{A}_1, \mathcal{A}_i, \mathcal{B}_1$ for the random variables denoting the respective outcomes and use the lowercase letters for their values; then, the joint distribution can be written as $p(\mathcal{A}_1 = a_1, \mathcal{A}_i = a_i, \mathcal{B}_1 = b_1|x{=}i, y{=}1)$.

for $i \in \{2, 3, \ldots, N_A\}$ and $j \in \{2, 3, \ldots, N_B\}$.

While these joint probability distributions are not empirically accessible[4], according to the experimental setup they must satisfy

$$p(a_1, b_1|x = 1, y = 1) = \wp(a_1, b_1), \tag{4.15a}$$

$$\sum_{a_1} p(a_1, a_i, b_1|x = i, y = 1) = \wp(a_i, b_1), \tag{4.15b}$$

$$\sum_{b_1} p(a_1, b_1, b_j|x = 1, y = j) = \wp(a_1, b_j), \tag{4.15c}$$

$$\sum_{a_1, b_1} p(a_1, a_i, b_1, b_j|x = i, y = j) = \wp(a_i, b_j). \tag{4.15d}$$

We will now apply the other Local Friendliness assumption, Local Agency to this setting.

Assumption 6 (Local Agency)**.** The only relevant events correlated with an intervention are in its future light cone.

Since in LF scenarios, we assume that none of a_i, a_1 or b_1 is in the future light cone of Bob's measurement setting y, and none of b_j, a_1 or b_1 is in the future light cone of Alice's measurement setting x, Local Agency implies that

$$p(a_1, b_1|x, y) = p(a_1, b_1) \tag{4.16a}$$

$$p(a_1, a_i, b_1|x = i, y) = p(a_1, a_i, b_1|x = i), \tag{4.16b}$$

$$p(a_1, b_1, b_j|x, y = j) = p(a_1, b_1, b_j|y = j). \tag{4.16c}$$

The existence of the distributions in Eq. (4.14) and the conditions in Eqs. (4.15) and (4.16) constitute the *LF constraints* on the empirical correlations in the LF scenario, leading to the so-called *LF inequalities*. The set of empirical correlations consistent with all LF inequalities is termed the *LF polytope*. The Local Friendliness no-go theorem [12, 13, 265], as we have seen in Section 2.5, then shows that some LF inequalities can be violated in a proposed quantum realization of the LF scenario. In the sections below, we will explore the details of LF inequalities for various LF scenarios, including generalizations to the ones defined in Section 2.5.1.

[4]For instance, the joint probability distribution $p(a_1, b_1, b_2|x = 1, y = 2)$ is not empirically accessible as for the setting $y = 2$ Alice reverses Debbie's measurement B_1, thereby erasing outcome b_1 and destroying every record of it. Consequently, for $y = 2$ no observer can jointly access records of both outcomes b_1 and b_2.

4.2.2 The compatibility graph for a N_A, N_B-setting LF scenario

We start our results by applying the notion of a compatibility graph introduced in Section 2.2.1 to the N_A, N_B-setting LF scenario defined in Section 2.5.1. In a N_A, N_B-setting LF scenario, all measurements done on Alice's side, i.e., $A_1, \ldots, A_{N_A}$, are compatible with all measurements done on Bob's side, i.e., $B_1, \ldots, B_{N_B}$. However, $A_1, \ldots, A_{N_A}$ measurements are not necessarily compatible with each other, and similarly, $B_1, \ldots, B_{N_B}$ measurements are not necessarily compatible with each other. Thus, the corresponding compatibility graph for the N_A, N_B-setting LF scenario is as shown in Fig. 4.3 where each of $A_1, \ldots, A_{N_A}$ is connected with each of $B_1, \ldots, B_{N_B}$, but no connections among $A_1, \ldots, A_{N_A}$ or among $B_1, \ldots, B_{N_B}$.

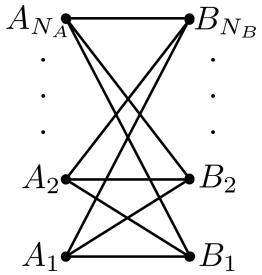

Figure 4.3: Compatibility graph for a N_A, N_B-setting LF scenario. Each node is associated with a measurement. The measurements connected by an edge are compatible and form a context. Every node on one side is connected by an edge with every node on the other side, forming a complete bipartite graph.

Such a compatibility graph is equivalent to a compatibility graph of a bipartite Bell scenario where in each run, Alice implements one of the measurements $A_1, \ldots, A_{N_A}$ while Bob implements one of the measurements $B_1, \ldots, B_{N_B}$. We refer to such a bipartite Bell scenario as *the corresponding Bell scenario for the N_A, N_B-setting LF scenario*.

Note that the constraints coming from Local Friendliness, presented in Section 4.2.1, imply the no-disturbance constraints (cf. Section 2.2.1) on the empirical correlations in a N_A, N_B-setting LF scenario. These no-disturbance constraints are

$$\sum_{b_l} \wp(a_k, b_l) = \sum_{b_{l'}} \wp(a_k, b_{l'}) \tag{4.17a}$$

$$\sum_{a_k} \wp(a_k, b_l) = \sum_{a_{k'}} \wp(a_{k'}, b_l), \tag{4.17b}$$

for all $k, k' \in \{1, 2, \ldots, N_A\}$ and $l, l' \in \{1, 2, \ldots, N_B\}$. They are obtained by taking the appropriate marginals of distributions in Eq. (4.16) and then connecting them

with empirical correlations via Eq. (4.15). For instance, to prove that $\sum_{b_1} \wp(a_1, b_1) = \sum_{b_2} \wp(a_1, b_2)$,

$$\begin{aligned}\sum_{b_1} \wp(a_1, b_1) &= \sum_{b_1} p(a_1, b_1|x = 1, y = 1) = \sum_{b_1} p(a_1, b_1|x = 1, y = 2) \\ &= \sum_{b_1,b_2} p(a_1, b_1, b_2|x = 1, y = 2) \\ &= \sum_{b_2} p(a_1, b_2|x = 1, y = 2) = \sum_{b_2} \wp(a_1, b_2)\end{aligned} \tag{4.18}$$

where the first and last equalities hold by Eq. (4.15) and the second and second-to-last by Eq. (4.16), and the third and fourth equalities by marginalization of probabilities. Furthermore, all (expected) empirical correlations in a quantum experimental realization of an LF scenario satisfy these no-disturbance relations (which can also be seen as no-signaling here).

As the LF constraints imply no-disturbance, it makes sense to further ask if LF constraints can be understood as KSNC constraints in some sense, which is what we do next in this section. In particular, we will show that the LF constraints on the empirical correlations in the LF scenario sometimes (and we will see when) coincide with the KSNC constraints on the corresponding compatibility graph for some N_A, N_B and always imply KSNC constraints on subgraphs of that graph. In the cases where they coincide (cf. Sections 4.2.4 and 4.2.5), the KSNC constraints on the corresponding compatibility graph provide all LF inequalities, and the LF polytope coincides with the Bell polytope of the corresponding Bell scenario. In other cases (cf. Section 4.2.6), the LF constraints on the empirical correlations in the LF scenario are *strictly weaker* than the KSNC constraints on the corresponding compatibility graph. In such scenarios, some (not all) LF inequalities can be derived via demanding KSNC constraints on *subgraphs* of the corresponding graph and the LF polytope strictly contains the Bell polytope. Finally, we generalize the LF scenario recapped in Section 4.2.1 to multipartite sequential LF scenarios and we characterize a wide class of such scenarios where all the LF inequalities are KSNC constraints on the corresponding graph, meaning that the LF polytope coincides with the Bell polytope for the corresponding multipartite Bell scenario.

Summarizing, we generalize Local Friendliness scenarios, point out connections between Local Friendliness and KS noncontextuality and thereby derive a wide class of Local Friendliness inequalities in these scenarios.

Before deriving these results, we will first provide an alternative expression in Section 4.2.3 of the LF constraints defined in Section 4.2.1, which will aid in our investigation

of the relationship between LF constraints and KSNC constraints.

4.2.3 An alternative expression of the LF constraints

We saw in Section 4.2.1 that the constraints implied by LF in the N_A, N_B-setting LF scenario are the existence of probability distributions as in Eq. (4.14) and satisfying Eqs. (4.15) and (4.16). Next, we introduce an alternative expression of these constraints; one which will later be useful for making the connection to KSNC. Two of its immediate consequences are shown in Section 4.2.3 and Section 4.2.3.

This alternative expression is the following. There must exist a joint distribution $p_{ij}(a_1, b_1, a_i, b_j)$ for any $i \in \{2, \ldots, N_A\}$, and $j \in \{2, \ldots, N_B\}$ with the following two properties: 1) each $p_{ij}(a_1, a_i, b_1, b_j)$ has as its marginals the four pairwise empirical correlations $\wp(a_i, b_j)$, $\wp(a_1, b_j)$, $\wp(a_i, b_1)$ and $\wp(a_1, b_1)$; 2) they agree on their overlaps, i.e.,

$$\sum_{a_i, b_j} p_{ij}(a_1, a_i, b_1, b_j) = \sum_{a_{i'}, b_{j'}} p_{i'j'}(a_1, a_{i'}, b_1, b_{j'}), \tag{4.19a}$$

$$\sum_{b_j} p_{ij}(a_1, a_i, b_1, b_j) = \sum_{b_{j'}} p_{ij'}(a_1, a_i, b_1, b_{j'}), \tag{4.19b}$$

$$\sum_{a_i} p_{ij}(a_1, a_i, b_1, b_j) = \sum_{a_{i'}} p_{i'j}(a_1, a_{i'}, b_1, b_j). \tag{4.19c}$$

We will now prove that this is equivalent to the LF constraints defined in Section 4.2.1.

First, we will show that the LF constraints imply the existence of the distributions $\{p_{ij}(a_1, b_1, a_i, b_j)\}_{i \in \{2,\ldots,N_A\}, j \in \{2,\ldots,N_B\}}$ with the two properties mentioned above.

The LF constraints include the existence of $p(a_1, a_i, b_1, b_j | x = i, y = j)$ for any $i \in \{2, 3, \ldots, N_A\}$ and $j \in \{2, 3, \ldots, N_B\}$ as per Eq. (4.14d), a consequence of the assumption of AOE in the N_A, N_B-setting LF scenario. Now, let

$$p_{ij}(a_1, a_i, b_1, b_j) = p(a_1, a_i, b_1, b_j | x = i, y = j). \tag{4.20}$$

Then, Eqs. (4.15) and (4.16) together imply that

$$\sum_{a_1, b_1} p_{ij}(a_1, a_i, b_1, b_j) = \wp(a_i, b_j), \tag{4.21a}$$

$$\sum_{a_i, b_1} p_{ij}(a_1, a_i, b_1, b_j) = \wp(a_1, b_j), \tag{4.21b}$$

$$\sum_{a_1, b_j} p_{ij}(a_1, a_i, b_1, b_j) = \wp(a_i, b_1), \tag{4.21c}$$

$$\sum_{a_i, b_j} p_{ij}(a_1, a_i, b_1, b_j) = \wp(a_1, b_1). \tag{4.21d}$$

Here, Eq. (4.21a) is equivalent to Eq. (4.15d) using Eq. (4.20), and Eq. (4.21b) comes from the fact that

$$\begin{aligned}
&\sum_{a_i,b_1} p_{ij}(a_1,a_i,b_1,b_j)\\
=&\sum_{a_i,b_1} p(a_1,a_i,b_1,b_j|x=i,y=j)\\
=&p(a_1,b_j|x=i,y=j)\\
=&p(a_1,b_j|x=1,y=j)\\
=&\wp(a_1,b_j),
\end{aligned} \tag{4.22}$$

where we used Eq. (4.16c) for the third equality and Eq. (4.15c) for the fourth equality; the proof for Eq. (4.21c) or Eq. (4.21d) is similar.

Eqs. (4.21) mean that the joint distribution $p_{ij}(a_1,a_i,b_1,b_j)$ has the four pairwise empirical correlations $\wp(a_i,b_j)$, $\wp(a_1,b_j)$, $\wp(a_i,b_1)$ and $\wp(a_1,b_1)$ as its marginals, proving the first property.

Furthermore, the LF constraints due to the assumption of Local Agency in Eqs. (4.16) directly imply Eqs. (4.19) by Eq. (4.20), meaning that the distributions in

$$\{p_{ij}(a_1,b_1,a_i,b_j)\}_{i\in\{2,\dots,N_A\},j\in\{2,\dots,N_B\}}$$

agree on their overlaps, proving the second property.

Now, we will prove the converse direction: that the existence of

$$\{p_{ij}(a_1,b_1,a_i,b_j)\}_{i\in\{2,\dots,N_A\},j\in\{2,\dots,N_B\}}$$

with the above mentioned two properties implies all LF constraints.

For any $i\in\{2,\dots,N_A\}$ and $j\in\{2,\dots,N_B\}$, define

$$p(a_1,b_1|x=1,y=1) \coloneqq \sum_{a_2,b_2} p_{22}(a_1,a_2,b_1,b_2) \tag{4.23a}$$

$$p(a_1,a_i,b_1|x=i,y=1) \coloneqq \sum_{b_2} p_{i2}(a_1,a_i,b_1,b_2) \tag{4.23b}$$

$$p(a_1,b_1,b_j|x=1,y=j) \coloneqq \sum_{a_2} p_{2j}(a_1,a_2,b_1,b_j) \tag{4.23c}$$

$$p(a_1,a_i,b_1,b_j|x=i,y=j) \coloneqq p_{ij}(a_1,a_i,b_1,b_j) \tag{4.23d}$$

These are the distributions required to exist by AOE in Eqs. (4.14). They reproduce the empirical correlations as in Eqs. (4.15) due to the first property of

$$\{p_{ij}(a_1,b_1,a_i,b_j)\}_{i\in\{2,\dots,N_A\},j\in\{2,\dots,N_B\}}$$

as in Eqs. (4.21).

Furthermore, the LF conditions in Eq. (4.16) are also satisfied due to the second property of $\{p_{ij}(a_1, b_1, a_i, b_j)\}_{i\in\{2,\ldots,N_A\},j\in\{2,\ldots,N_B\}}$ that these distributions agree on their overlaps. For example, to prove Eq. (4.16b), it suffices to note that for any $i \in \{2, \ldots, N_A\}$, $j \in \{2, \ldots, N_B\}$ and $j' \in \{1, 2, \ldots, N_B\}$, we have

$$\begin{aligned} & p(a_1, a_i, b_1|x = i, y = j) \\ =& \sum_{b_j} p_{ij}(a_1, a_i, b_1, b_j) \\ =& \sum_{b_{j'}} p_{ij'}(a_1, a_i, b_1, b'_j) \\ =& p(a_1, a_i, b_1|x = i, y = j'), \end{aligned} \tag{4.24}$$

where the second equality is due to the fact that $p_{ij}(a_1, a_i, b_1, b_j)$ and $p_{ij'}(a_1, a_i, b_1, b'_j)$ agree on their overlap.

By expressing the LF constraints in terms of the existence of distributions in $\{p_{ij}(a_1, b_1, a_i, b_j)\}_{i\in\{2,\ldots,N_A\},j\in\{2,\ldots,N_B\}}$ and the two properties they have, we can easily obtain two insights into the features of the LF inequalities, as we now show.

4.2.3.1 *A class of tight LF inequalities that are KSNC constraints on subgraphs*

The first insight that immediately follows from the alternative expression just derived is the existence of a class of LF inequalities that are KSNC constraints on subgraphs of the compatibility graph for a N_A, N_B-setting LF scenario, i.e., Fig. 4.3. Specifically, all the LF inequalities in this class are KSNC constraints on graphs of the form shown in Fig. 4.4.

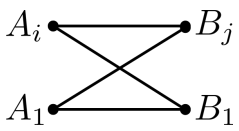

Figure 4.4: The compatibility graph for A_1, A_i, B_1 and B_j with $i \in \{2, 3, \ldots, N_A\}$ and $j \in \{2, 3, \ldots, N_B\}$. It is a subgraph of Fig. 4.3.

To see this, first recall from Section 2.2.1 that demanding KSNC constraints on a compatibility graph is equivalent to demanding the existence of a joint probability distribution over the outcomes of all measurements associated with vertices of the graph, such that the distribution has the empirical correlation associated with each edge as its marginals. For the compatibility graph in Fig. 4.4, from Eq. (4.21), we see that

$p_{ij}(a_1, a_i, b_1, b_j)$ is a joint distribution over the outcomes of the four measurements in the graph and has as its marginals the four empirical correlations for the four respective edges, namely, $\wp(a_i, b_j)$, $\wp(a_1, b_j)$, $\wp(a_i, b_1)$ and $\wp(a_1, b_1)$.

Thus, the LF constraints require the KSNC constraints on graphs in the form of the one in Fig. 4.4, which are subgraphs of Fig. 4.3, the full compatibility graph for a N_A, N_B-setting LF scenario. Furthermore, since Fig. 4.3 corresponds to a bipartite Bell scenario where both Alice and Bob have binary measurement settings, the KSNC constraints on Fig. 4.3 are CHSH inequalities. Therefore, for each $i \in \{2, 3, \ldots, N_A\}$ and $j \in \{2, 3, \ldots, N_B\}$, we can derive LF inequalities that are CHSH inequalities on $\wp(a_i, b_j)$, $\wp(a_1, b_j)$, $\wp(a_i, b_1)$ and $\wp(a_1, b_1)$.

Furthermore, these LF inequalities are in fact *tight*, meaning that they can be saturated by correlations that satisfy the LF assumptions. This is because of the second immediate insight we gain from the alternative expression of the LF constraints, which, as we will explain next, is that the LF inequalities are not stricter than the Bell inequalities for the corresponding Bell scenario, or in other words, the LF polytope always contains the corresponding Bell polytope. This fact has been previously shown in Refs. [12, 13, 17, 33] (and indirectly also in Ref. [341].)

4.2.3.2 *The LF polytope contains the Bell polytope*

Here we give an alternative proof of the fact that the LF polytope contains the Bell polytope in the corresponding Bell scenario to the ones in previous literature such as Refs. [12, 33].

The Bell polytope in the Bell scenario corresponding to our LF scenario is precisely characterized by the KSNC constraints on the compatibility graph of the LF scenario as in Section 4.2.2. That is, it comes from the requirement that there must exist a global distribution $p(a_1, b_1, a_2, b_2, \ldots, a_{N_A}, b_{N_B})$ over the outcomes of all measurements that reproduces all empirical correlations, namely, all $\wp(a_x, b_y)$ with $x \in \{1, 2, \ldots, N_A\}$ and $y \in \{1, 2, \ldots, N_A\}$.

When such a $p(a_1, b_1, a_2, b_2, \ldots, a_{N_A}, b_{N_B})$ exists, it further implies the existence of all distributions in $\{p_{ij}(a_1, b_1, a_i, b_j)\}_{i\in\{2,\ldots,N_A\}, j\in\{2,\ldots,N_B\}}$ with the two properties described at the beginning of Section 4.2.3. Specifically, we can define $p_{ij}(a_1, b_1, a_i, b_j)$ simply by marginalising $p(a_1, b_1, a_2, b_2, \ldots, a_{N_A}, b_{N_B})$. These distributions evidently reproduce the four empirical correlations $\wp(a_i, b_j)$, $\wp(a_1, b_j)$, $\wp(a_i, b_1)$ and $\wp(a_1, b_1)$; they also agree on their overlap by construction (as they all stem from the same global distribution).

Thus, the KSNC constraints (which are also Bell inequalities) for the corresponding Bell scenario imply the respective LF constraints, which also means that the LF polytope always contains the corresponding Bell polytope.

4.2.4 The 2,2-setting LF scenario

In this scenario, Alice and Bob have binary measurement settings. Thus, in total, we have four measurements A_1, A_2, B_1 and B_2 with measurement outcomes a_1, a_2, b_1 and b_2. The compatibility graph for the 2,2-setting LF scenario is shown in Fig. 4.5.

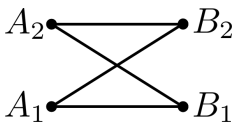

Figure 4.5: The compatibility graph for the 2,2-setting LF scenario, which is a special case of the graph in Fig. 4.3.

From the alternative expression described in Section 4.2.3 of the LF constraints, we know that there exists a joint distribution $p_{22}(a_1, a_2, b_1, b_2)$ such that it has as its marginals $\wp(a_2, b_2)$, $\wp(a_1, b_2)$, $\wp(a_2, b_1)$ and $\wp(a_1, b_1)$, which are all the empirical correlations in the 2,2-setting LF scenario. Thus, as discussed in Section 4.2.3, the four pairwise correlations $\wp(a_x, b_y)$ for $x = 1, 2$ and $y = 1, 2$ must obey KSNC constraints on the compatibility graph in Fig. 4.5, which give rise to the CHSH inequalities on these four correlations. Furthermore, since the LF polytope must contain the corresponding Bell polytope, as was proven in Section 4.2.3, in this 2,2-setting scenario, the LF polytope coincides with the Bell polytope whose nontrivial facets are characterized by these CHSH inequalities.

4.2.5 The 3,2-setting LF scenario

Now consider the case where Alice has 3 settings and Bob has 2 settings. There are now six empirical pairwise correlations, namely $\wp(a_x, b_y)$ for $x = 1, 2, 3$ and $y = 1, 2$.

Assuming Local Friendliness, following Section 4.2.3, we have that

- $p_{22}(a_1, a_2, b_1, b_2)$ has as its marginals $\wp(a_2, b_2)$, $\wp(a_1, b_2)$, $\wp(a_2, b_1)$ and $\wp(a_1, b_1)$. Thus, the four pairwise correlations $\wp(a_x, b_y)$ for $x = 1, 2$ and $y = 1, 2$ must obey KSNC constraints on the compatibility graph in Fig. 4.6b.

- $p_{32}(a_1, a_3, b_1, b_2)$ has as its marginals $\wp(a_3, b_2)$, $\wp(a_1, b_2)$, $\wp(a_3, b_1)$ and $\wp(a_1, b_1)$. Thus, the four pairwise correlations $\wp(a_x, b_y)$ for $x = 1, 3$ and $y = 1, 2$ must obey KSNC constraints on the compatibility graph in Fig. 4.6c.

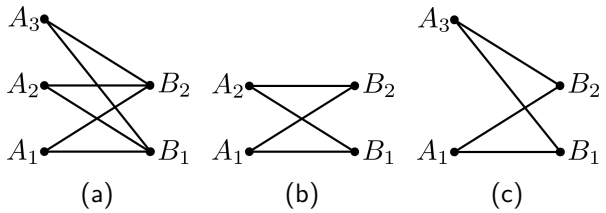

Figure 4.6: (a): The compatibility graph of the 3,2-setting LF scenario. (b) and (c): Subgraphs of the graph in (a) that are special cases of the graph in Fig. 4.4. In Section 4.2.5 we explain how LF assumptions allow us to 'glue' the distributions for (b) and (c) together to form a joint distribution for (a), using the fact that LF requires the distributions for (b) and (c) to agree on their overlap, i.e., have the same marginals on the intersection of their domain.

So far, we have *not* shown that the four pairwise correlations $\wp(a_x, b_y)$ for $x = 2, 3$ and $y = 1, 2$ must satisfy KSNC constraints on a corresponding compatibility graph, and in fact one cannot do so by an argument analogous to the above.

Nevertheless, we will now show that the Local Friendliness assumptions do require KSNC constraints for these correlations as well, and in fact, that the Local Friendliness assumptions imply KSNC constraints on the compatibility graph for the whole 3,2-LF scenario as shown in Fig. 4.6a, meaning that all six pairwise correlations $\wp(a_x, b_y)$ for $x = 1, 2, 3$ and $y = 1, 2$ are marginals of a single joint probability distribution.

4.2.5.1 *The 'gluing' trick*

The aforementioned joint distribution of which all six pairwise correlations $\wp(a_x, b_y)$ for $x = 1, 2, 3$ and $y = 1, 2$ are marginals can be constructed using $p_{22}(a_1, a_2, b_1, b_2)$ and $p_{32}(a_1, a_3, b_1, b_2)$ via

$$p_{22,32} := \frac{p_{22}(a_1, a_2, b_1, b_2)p_{32}(a_1, a_3, b_1, b_2)}{\sum_{a_3} p_{32}(a_1, a_3, b_1, b_2)}. \tag{4.25}$$

We call the trick used in Eq. (4.25) for constructing $p_{22,32}$ the *gluing* operation[5], which constructs a larger joint probability distribution (i.e., $p_{22,32}$) from smaller distributions (i.e., $p_{22}(a_1, a_2, b_1, b_2)$ and $p_{32}(a_1, a_3, b_1, b_2)$) in such a way that (as we will prove soon) the larger one has as its marginals all empirical pairwise correlations that are marginals of the smaller ones. We call it 'gluing' since in terms of the compatibility graphs upon

[5]A similar technique was used in, e.g., Refs. [17, 342].

which the KSNC constraints are demanded by the respective joint distributions, such a procedure 'glues' together two smaller graphs (i.e., Fig. 4.6b and Fig. 4.6c) to a bigger one (i.e., Fig. 4.6a); in terms of the probability distributions, it 'glues' together two marginal distributions to create a bigger joint distribution.

The proof for $p_{22,32}$ to recover all six pairwise correlations uses the fact that $p_{22}(a_1, a_2, b_1, b_2)$ and $p_{32}(a_1, a_3, b_1, b_2)$ agree on their overlap:

$$\sum_{a_2} p_{22}(a_1, a_2, b_1, b_2) = \sum_{a_3} p_{32}(a_1, a_3, b_1, b_2), \tag{4.26}$$

which is a special case of Eq. (4.19c). See Appendix B of Ref. [21] for the explicit proof.

As such, all six pairwise empirical correlations in the 3,2-setting LF scenario can be recovered as marginals of a single joint distribution $p_{22,32}$. Thus, these six pairwise correlations must satisfy the KSNC constraints on the compatibility graph in Fig. 4.6a. As this compatibility graph also represents a Bell scenario, it follows that these constraints give rise to Bell inequalities. Furthermore, these Bell inequalities are tight LF inequalities since the LF polytope must contain the Bell polytope as proven in Section 4.2.3. Since these six pairwise correlations are all the observable correlations in the 3,2-setting LF scenario, then, the Local Friendliness polytope coincides with the Bell polytope in this case.

A trivial corollary of the above is that the four pairwise correlations $\wp(a_x, b_y)$ for $x = 2, 3$ and $y = 1, 2$ can be recovered as marginals of a single joint distribution, namely, $p_{22,32}$, and so must satisfy CHSH inequalities, and equivalently, the constraints implied by KSNC for any four projective measurements satisfying the compatibility relations in Fig. 4.4 (with the appropriate relabeling).

Following an analogous proof to the one we just did here for the 3,2-setting LF scenario, one can further show that a $(N_A, 2)$-setting or a $(2, N_B)$-setting LF scenario for any integer $N_A, N_B \geq 2$ has its LF constraints being equivalent to the KSNC constraints on the compatibility graph for that LF scenario, which give rise to the Bell inequalities for the corresponding Bell scenario.

4.2.5.2 The minimal LF scenario

An analogous argument can be constructed to explain why the LF polytope for the so-called 'minimal LF scenario' studied in Ref. [33, 265] also coincides with the corresponding Bell polytope.

In the minimal LF scenario, there is no Debbie. Alice has a binary measurement setting. She and Charlie proceed as usual. Bob, however, now chooses between two

possible measurements B_y on the other half of the bipartite system depending on his binary setting choice y. The empirical correlations in this scenario are the four pairwise correlations: $\wp(a_2, b_2)$, $\wp(a_1, b_2)$, $\wp(a_2, b_1)$ and $\wp(a_1, b_1)$.

Here, AOE demands the existence of $p(a_1, b_1|x = 1, y = 1)$, $p(a_1a_2b_1|x = 2, y = 1)$ and $p(a_1a_2b_2|x = 2, y = 2)$ for the corresponding runs of the experiment. As shown in Appendix B of Ref. [21], by applying the gluing trick to $p(a_1a_2b_1|x = 2, y = 1)$ and $p(a_1a_2b_2|x = 2, y = 2)$, we obtain a joint probability distribution that recovers all four pairwise correlations, and thus explains why the LF polytope coincides with the corresponding Bell polytope in the minimal scenario.

Furthermore, one can generalize the minimal LF scenario and show that if Bob's measurement choice is not binary but instead with higher cardinality for any integer $N_B \geq 2$, it is still true that all LF constraints in this scenario are equivalent to the KSNC constraints (or equivalently the Bell inequalities) on a corresponding Bell scenario.

4.2.6 The 3,3-setting LF scenario

Now consider the case where both Alice and Bob have 3 settings. In this case, the Local Friendliness polytope does *not* coincide with the Bell polytope, as was shown in Ref. [12], which used a software to derive the LF polytope for binary outcomes explicitly. We will give a more algebraic derivation and explanation of this fact (and without the restriction to binary outcomes).

Assuming Local Friendliness, following Section 4.2.3, we have that $p_{22}(a_1, a_2, b_1, b_2)$ has as its marginals $\wp(a_2, b_2)$, $\wp(a_1, b_2)$, $\wp(a_2, b_1)$ and $\wp(a_1, b_1)$. Thus, the four pairwise correlations $\wp(a_x, b_y)$ for $x = 1, 2$ and $y = 1, 2$ must obey the KSNC constraints on the compatibility relations in Fig. 4.7a. Since this compatibility graph in Fig. 4.7a corresponds to a Bell scenario, these KSNC constraints give rise to Bell inequalities. Similar reasoning holds for $p_{32}(a_1, a_3, b_1, b_2)$, $p_{23}(a_1, a_2, b_1, b_3)$ and $p_{33}(a_1, a_3, b_1, b_3)$, depicted in Figs. 4.7b to 4.7d.

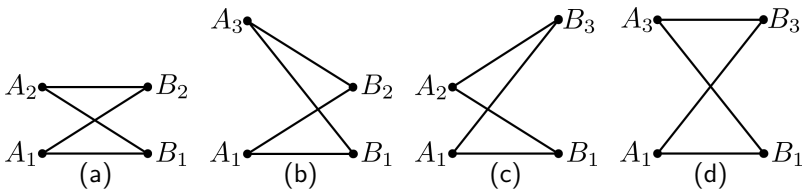

Figure 4.7: Special cases of the compatibility graph in Fig. 4.3 that are subgraphs of the compatibility graph of the 3,3-setting LF scenario.

Furthermore, as proven in Section 4.2.3, these four distributions, $p_{22}, p_{23}, p_{32}, p_{33}$ (where we omit their arguments for compactness), agree on their overlaps. Then, by gluing various three out of the four distributions together, following an argument like that for Eq. (4.25), we obtain the following. Gluing together p_{22}, p_{32}, p_{23} we find that $p_{22,32,23} \coloneqq \frac{p_{22}p_{32}p_{23}}{\sum_{a_3} p_{32} \sum_{b_3} p_{23}}$ has as its marginals all pairwise correlations $\wp(a_x, b_y)$ for $x = 1, 2, 3$ and $y = 1, 2, 3$ *except* for $\wp(a_3, b_3)$, and thus, we also denote $p_{22,32,23}$ as $p_{\neg a_3 b_3}$. As such, these pairwise correlations must obey the KSNC constraints on the compatibility graph in Fig. 4.8a. The same reasoning holds for $p_{\neg a_2 b_2}, p_{\neg a_2 b_3}, p_{\neg a_3 b_2}$ with compatibility graphs in Figs. 4.8b to 4.8d.

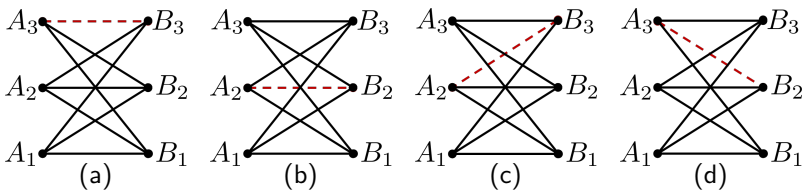

Figure 4.8: Subgraphs of the compatibility graph of the 3,3-setting LF scenario. Each of them has 8 edges, i.e., 8 pairwise compatibility relations. The red dashed line indicates the missing compatibility relation for deriving Bell inequalities on all six measurements in each graph.

Unlike all of the previous cases, these compatibility graphs in Fig. 4.8 are *not* the complete compatibility graphs of any Bell scenario, because of the missing edges indicated by the red dashed lines in Fig. 4.8; rather, they are subgraphs. Consequently, the inequalities derived from KSNC constraints in this case will not generally be tight Bell inequalities. Nevertheless, some of these KSNC inequalities coincide with tight Bell inequalities because the compatibility graph in Fig. 4.8 includes subgraphs that are the complete compatibility graphs for some other Bell scenarios. For example, Fig. 4.8a includes the graphs in Figs. 4.7a to 4.7c and Fig. 4.6a as subgraphs, and they are complete compatibility graphs for some Bell scenarios; in addition, it includes a compatibility subgraph for $\{A_2, A_3, B_1, B_3\}$, one for $\{A_1, A_3, B_2, B_3\}$, one for $\{A_2, A_3, B_1, B_2\}$, one for $\{A_1, A_2, B_2, B_3\}$ and one for $\{A_2, A_3, B_1, B_2, B_3\}$, of which each represents a complete compatibility graph for a Bell scenario. Thus, the KSNC constraints on the compatibility graphs in Fig. 4.8 include Bell inequalities for these compatibility subgraphs.

The compatibility graphs in Fig. 4.8 are also *not* the compatibility graph of the 3,3-setting LF scenario because of the missing edges depicted as red dashed lines. In fact, the Local Friendliness assumptions do not imply the existence of a joint probability distribution that can have as its marginals all the 9 pairwise empirical correlations $\wp(a_x, b_y)$ for $x = 1, 2, 3$ and $y = 1, 2, 3$ in the 3,3-setting LF scenario, and hence, the LF polytope

here is strictly bigger than the Bell polytope, as pointed out in Ref. [12]. In Appendix B of Ref. [21], we provide further intuitions on why this is the case by considering how the gluing trick introduced in Section 4.2.5 works.

Since any (N_A, N_B)-setting LF scenario with $N_A, N_B \geq 3$ has the 3,3-setting LF scenario as a subscenario, we can further see that a (N_A, N_B)-setting LF scenario for any integer $N_A, N_B \geq 3$ will have its LF constraints being strictly weaker than the KSNC constraints on the compatibility graph for that LF scenario, meaning that the LF polytope is strictly bigger than the corresponding Bell polytope.

Note that the KSNC constraints on the four graphs in Fig. 4.8 do not have all constraints given by Local Agency. In particular, Local Agency further demands certain constraints on the relationship between $p_{\neg a_3 b_3}$, $p_{\neg a_2 b_2}$, $p_{\neg a_2 b_3}$ and $p_{\neg a_3 b_2}$, due to the fact that any two of the building blocks of the gluing trick (namely, p_{22}, p_{23}, p_{32} and p_{33}) must agree on their overlap. Thus, the inequalities we have derived here are not necessarily tight. This can also be seen by comparing to the results in Ref. [12] (where tight inequalities were derived numerically for a specific scenario); see Appendix C of Ref. [21] for more details.

Although the inequalities we have just derived are not necessarily tight (such as some of the ones in Section 4.2.6), and so do not fully characterize the set of correlations satisfying LF assumptions, they are nonetheless tight enough to admit violations within unitary quantum theory. Every LF inequality (whether it is tight or not) is also a valid Bell inequality, as a consequence of the fact that the LF polytope always contains the Bell polytope for the same setting and outcome variable cardinalities. If such a Bell inequality can be violated by quantum theory in a Bell scenario, then it, as a LF inequality, can also be violated by quantum theory. This is because one can translate any quantum proposal for a N_A, N_B-setting Bell scenario violating a Bell inequality to a quantum proposal for the N_A, N_B-setting Local Friendliness scenario that violates the corresponding LF inequality. See Appendix D of Ref. [21] for more details.

4.2.7 Generalizations to multipartite sequential LF scenarios

So far, we have considered bipartite LF scenarios and characterized the relationship between the LF polytope in these scenarios and their corresponding KSNC or Bell polytopes. This analysis can be extended to more general LF scenarios involving multiple superobserver–observer pairs performing measurements sequentially (with arbitrarily

long sequences). Furthermore, we can characterize the conditions under which all LF constraints in such a scenario can also be interpreted as KSNC constraints, and in particular, as Bell inequalities. We refer interested readers to Appendix E of Ref. [21] for more details.

4.3 Constructing new LF scenarios from KS noncontextuality arguments

In Section 4.2, we focused on showing how, given a specific LF scenario, one can derive many LF inequalities using known tools and results from the literature on KSNC, and on clarifying the relation between Bell local causality and LF. In this section, we show how given a possibilistic proof of the failure of KSNC one can construct an extended Wigner's friend argument—a no-go theorem against the Local Friendliness assumptions of AOE and Local Agency.

Compared to the measurements of the underlying KSNC arguments, some additional measurements are introduced. Without these additional measurements, no-go theorems against AOE can be formulated, but only together with the assumption of Commutation Irrelevance of Section 4.1 replacing Local Agency. Details can be found in Appendix K of Ref. [21].

4.3.1 The 5-cycle example

In Section 4.1, it was shown that the 5-cycle noncontextuality scenario [237, 335, 336] can be used to construct an extended Wigner's friend argument with the assumption of Commutation Irrelevance, which is not one of the LF assumptions. Here, we show that the protocol of Section 4.1 can in fact be adapted to construct a novel scenario where one can prove the LF no-go theorem [12] (cf. Theorem 1), that is, a no-go theorem based on the LF assumptions without using Commutation Irrelevance.

4.3.1.1 The LF 5-cycle protocol

We leverage the 5-cycle KSNC argument of Section 4.1.2 to construct a novel scenario in which one can prove the LF no-go theorem. Our presentation of this LF 5-cycle protocol highlights the differences from the protocol in Section 4.1.

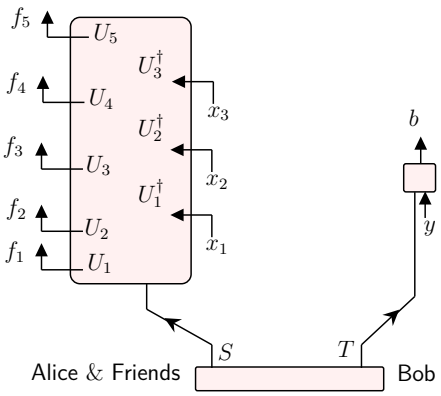

Figure 4.9: The schematic representation of the LF 5-cycle scenario, see also Figure 4.10.

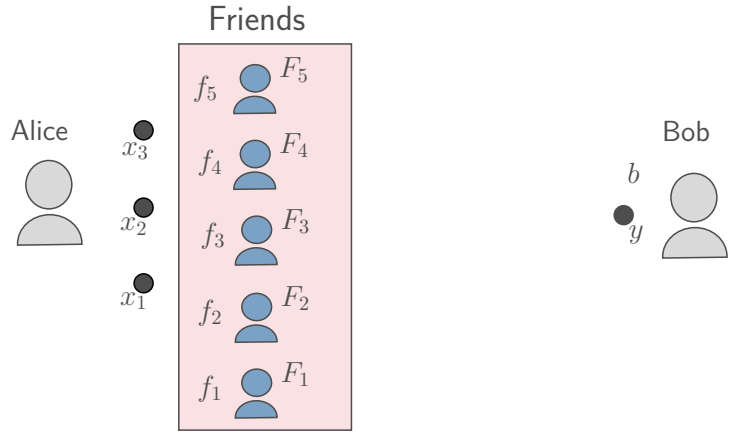

Figure 4.10: Cartoon representation of the 5-cycle LF setup, with Alice a superobserver for F_1, F_2, F_3, F_4, F_5, who perform measurements. Alice makes choices x_1, x_2, x_3 whether to reveal measurement outcomes and Bob, spacelike separated, makes measurement choice y and obtains outcome b.

This new LF scenario involves two groups of agents. The first group consists of Alice (denoted A) and five friends (denoted F_1, ..., F_5). The second group consists of a single agent Bob. See Figure 4.10 for a cartoon representation of the set-up. Compared to the protocol of Section 4.1, we have one more agent here, namely, Bob; furthermore, the superobserver Alice here has measurement choices. These additions are crucial for deriving a no-go theorem based on LF assumptions without using the assumption of Commutation Irrelevance of Section 4.1.

In the following, we will describe the quantum protocol that will be used to prove the no-go theorem. However, the scenario could also be described in a theory-independent way as discussed in Appendix A of Ref. [21].

At the beginning of the experiment, F_1 and Bob share a bipartite system ST where S

is a qutrit and T is a qubit. The joint system ST is prepared in the following entangled state:[6]

$$|\Psi\rangle_{ST} = \Pi_{v_1^\perp} |\eta\rangle_S \otimes |0\rangle_T + \langle v_1|\eta\rangle \, |v_1\rangle_S \otimes |1\rangle_T \tag{4.27}$$

where $\Pi_{v_1^\perp} := \mathbf{1} - |v_1\rangle\langle v_1|$. System S is sent to Alice's side while system T is sent to Bob.

On Alice's side, the protocol is similar to the one of Section 4.1. Alice has five friends, F_1, ..., F_5, who perform the five measurements A_i defined in Eq. (4.5) sequentially on the system S, with the superobserver's reversal of their measurements in between, i.e. applying the inverse unitary, as shown in Fig. 4.9. From AOE, each friend F_i observes an absolute outcome, denoted f_i, during their measurement A_i. We label the $|v_i\rangle\langle v_i|$ outcome by $f_i = 1$ and the other outcome by $f_i = 0$. In addition, the superobserver Alice has a binary choice x_i right before she implements the inverse unitary $U_i^\dagger$ for $i = 1, 2, 3$. If $x_i = 1$, Alice opens the door to F_i's and F_{i+1}'s labs and reveals their respective measurement outcome f_i and f_{i+1} to the world; if $x_i = 2$, she simply continues the protocol.[7] When $x_1 = x_2 = x_3 = 2$, the f_4 and f_5 outcomes are revealed to the world at the end of the experiment.

On Bob's side, Bob has a binary choice y. If $y = 1$, he makes a measurement $B_{0/1}$ on system T, which is a computational basis measurement; if $y = 2$, he makes a measurement $B_\pm$ on system T, which is a $\pm$ basis measurement, i.e, $\{|+\rangle\langle+|, |-\rangle\langle-|\}$. We denote Bob's absolute outcome by b.

[6] One way to construct this entangled state (4.27) is to let F_1 perform measurement A_1 on $|\eta\rangle_S$ and prepare $|0\rangle_T$ if her measurement outcome corresponds to $\Pi_{v_1^\perp}$ or $|1\rangle_T$ if her measurement outcome corresponds to $|v_1\rangle$, and then F_1 sends system T to Bob. In fact, instead of starting with F_1 and B sharing the entangled state (4.27), the protocol could start with F_1 having system S prepared in $|\eta\rangle_S$, followed by the above measure-prepare procedure. All subsequent operations by Alice, the friends F_i and Bob remain the same. In this alternative protocol, F_1's operation is in the past-light-cone of everyone else's operations, while Bob's operations are still space-like separated from all operations on Alice's side except for F_1's. This is similar to how the measure-prepare version of the Frauchiger–Renner argument [11] as it was originally presented, is equivalent to the simpler entanglement-based scenario (see, e.g., Refs. [24, 27]).

[7] Note that in the $x_i = 2$ case, all records of the outcome f_i are erased by the inverse unitaries applied by Alice; in contrast, in the $x_i = 1$ case, the act of revealing the outcome f_i to the world implies making and distributing many classical copies of it. In that case, the inverse unitaries performed by Alice do not have the effect of reverting the state of S and the Friends to their initial states. They moreover will strongly impact outcomes observed by friends later in the protocol. However, all this is irrelevant, because in those cases the later outcomes of the friends are not used in the proof of the LF no-go theorem. Indeed, one could avoid these subtleties simply by stipulating that when $x_i = 1$, Alice not only reveals the two immediately prior measurement outcomes to the world but also terminates the protocol.

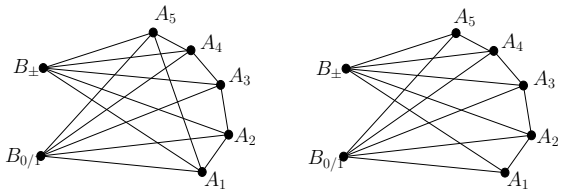

Figure 4.11: Left: A compatibility graph for the measurements in the LF 5-cycle protocol of Section 4.3.1. Right: The constraints from LF assumptions do not impose KSNC relative to the compatibility graph on the left, but they do effectively impose the same constraints as KSNC relative to the subgraph obtained when the compatibility relation between A_1 and A_5 is absent, as shown on the right. (They may impose further constraints in addition to these.)

We assume that all operations on Alice's side are done at space-like separation from Bob's measurement setting and outcome.

One can naturally associate a compatibility graph with this 5-cycle LF scenario: namely the one shown in the left subfigure of Fig. 4.11. The Local Friendliness assumptions on this LF 5-cycle scenario *do not* imply KSNC on this compatibility graph, nor on the compatibility graph for the 5-cycle scenario (shown in Fig. 4.1), since the joint distribution over A_1 and A_5 measurement outcomes is not observable. Rather, the LF assumptions imply KSNC constraints on the graph in the right figure of Fig. 4.11 (where the edge between A_1 and A_5 is absent). Nevertheless, as we show now, the possibilistic proof of the failure of KSNC in the 5-cycle scenario can still be mathematically turned into a proof of the Local Friendliness no-go theorem.

4.3.1.2 The no-go theorem

A schematic of the proof is given in Figure 4.12. The proof begins by noting that AOE demands the existence of the conditional probability distribution

$$p(f_1, f_2, f_3, f_4, f_5, b|x_1, x_2, x_3, y), \tag{4.28}$$

exists (although of course, it is not accessible to any agent). Next, one considers the predictions made by quantum theory for various *empirical* correlations in the LF 5-cycle protocol, and then applies Local Agency to them to obtain that

$$p(f_1 = 1, f_2 = 1|x_{1-3} = 2, y = 2, b = +) = 0, \tag{4.29a}$$
$$p(f_2 = 0, f_3 = 0|x_{1-3} = 2, y = 2, b = +) = 0, \tag{4.29b}$$
$$P(f_3 = 1, f_4 = 1|x_{1-3} = 2, y = 2, b = +) = 0, \tag{4.29c}$$
$$P(f_4 = 0, f_5 = 0|x_{1-3} = 2, y = 2, b = +) = 0, \tag{4.29d}$$

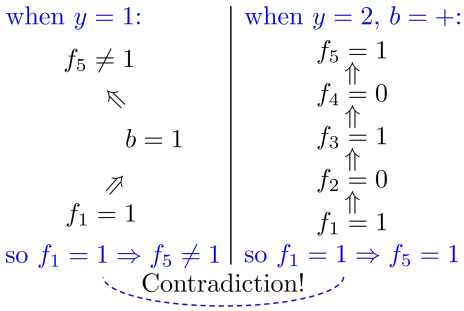

Figure 4.12: The schematic proof using the LF 5-cycle protocol for the LF no-go theorem. The key steps in the proof are as follows. When Bob measures his qubit in the computational basis (i.e., when $y = 1$, as in the left side of the figure), his outcome b must be the same as the f_1 outcome, which in turn constrains the correlation over f_1 and f_5. When Bob measures his qubit in the $\pm$ basis (i.e., when $y = 2$) and obtains outcome $b = +$ (as in the right side of the figure), the effective state on Alice and her friends' side becomes $|\eta\rangle_S$, i.e., the state needed in the 5-cycle proof of KSNC theorem, and one obtains a contradictory constraint on the correlation between f_1 and f_5 under the Local Friendliness assumptions as Local Agency requires the correlation f_1, f_5 to be independent of y.

and

$$p(f_1 = 1, f_5 = 1 | x_{1-3} = 2, y = 2, b = +) = 0. \tag{4.30}$$

(A derivation of these equations is given below.) These five equations constitute constraints on five marginals of the joint distribution

$$p(f_1, f_2, f_3, f_4, f_5 | x_{1-3} = 2, y = 2, b = +), \tag{4.31}$$

which is well-defined due to the existence of Eq. (4.28) and the Born rule prediction that

$$\wp(b = + | x_{1-3} = 2, y = 2) > 0 \tag{4.32}$$

(a prediction that is empirically accessible).

However, these five constraints, i.e., Eqs. (4.29a) to (4.29d) and (4.30), cannot be simultaneously satisfied. This fact is the basis of the KSNC proof in the 5-cycle scenario; one can see it most easily by noting that these equations imply (respectively) that when $x_{1-3} = 2, y = 2, b = +$, we have

$$f_1 = 1 \Rightarrow f_2 = 0 \Rightarrow f_3 = 1 \Rightarrow f_4 = 0 \Rightarrow f_5 = 1. \tag{4.33}$$

and

$$f_1 = 1 \Rightarrow f_5 \neq 1, \tag{4.34}$$

which are in contradiction. Therefore there can be no joint distribution of the form in Eq. (4.30) consistent with the empirical predictions given by the Born rule and Local Agency. Hence the no-go theorem.

It remains only to establish Eqs. (4.29a)-(4.30).

We begin with Eqs. (4.29a) to (4.29d). Consider the following empirical predictions given by the Born rule for the cases where Bob measures in $\pm$ basis and obtains outcome $+$:

$$\wp(f_1 = 1, f_2 = 1 | x_1 = 1, x_2, x_3, y = 2, b = +) = 0, \tag{4.35a}$$
$$\wp(f_2 = 0, f_3 = 0 | x_1 = 2, x_2 = 1, x_3, y = 2, b = +) = 0, \tag{4.35b}$$
$$\wp(f_3 = 1, f_4 = 1 | x_{1-2} = 2, x_3 = 1, y = 2, b = +) = 0, \tag{4.35c}$$
$$\wp(f_4 = 0, f_5 = 0 | x_{1-3} = 2, y = 2, b = +) = 0. \tag{4.35d}$$

These are analogous to Eqs. (4.1a) to (4.1d); indeed, a quick way to see that these are the Born rule predictions in the present scenario is to realize that when $y = 2$ and $b = +$, the state S on Alice's side is effectively $|\eta\rangle$ defined in Eq. (4.4).

For Eq. (4.35a), since both f_1 and f_2 are in the past light cone of all Alice's choices, they are not in the future light cone of x_1, x_2, x_3 and according to Local Agency, we have

$$\begin{aligned} &\wp(f_1, f_2 | x_1 = 1, x_2, x_3, y = 2, b = +) \\ =&p(f_1, f_2 | x_{1-3} = 2, y = 2, b = +) \end{aligned} \tag{4.36}$$

Similarly, for Eqs. (4.35b) and (4.35c), Local Agency demands that

$$\begin{aligned} &\wp(f_2, f_3 | x_1 = 2, x_2 = 1, x_3, y = 2, b = +) \\ =&p(f_2, f_3 | x_{1-3} = 2, y = 2, b = +), \end{aligned} \tag{4.37}$$
$$\begin{aligned} &\wp(f_3, f_4 | x_{1-2} = 2, x_3 = 1, y = 2, b = +) \\ =&p(f_3, f_4 | x_{1-3} = 2, y = 2, b = +) \end{aligned} \tag{4.38}$$

Finally, Eq. (4.35d) is identical to Eq. (4.29d).

Therefore, together with Eqs. (4.35a) to (4.35d), one obtains Eqs. (4.29a) to (4.29d).

Now, we prove Eq. (4.30). To do so, we start with the following empirical predictions by the Born rule

$$\wp(f_1 = 1, b = 0 | x_1 = 1, x_2, x_3, y = 1) = 0, \tag{4.39}$$
$$\wp(f_5 = 1, b = 1 | x_{1-3} = 2, y = 1) = 0. \tag{4.40}$$

Eq. (4.39) can be seen by noticing the perfect correlation between the outcome of A_1 on system S and the outcome of computational basis measurement on system T given by the initial state $|\Psi_{ST}\rangle$. Eq. (4.40) can be seen by noticing that when the outcome on T corresponds to $|1\rangle_T$, the state of S is effectively $|v_1\rangle_S$, which is orthogonal to $|v_5\rangle_S$.

To relate these two empirical correlations, we use Local Agency again: Since neither f_1 nor b is in the future light cone of x_1, x_2, x_3, Local Agency demands that Eq. (4.39) satisfies

$$\begin{aligned}&\wp(f_1 = 1, b = 0|x_1 = 1, x_2, x_3, y = 1)\\ =&p(f_1 = 1, b = 0|x_{1-3} = 2, y = 1) = 0.\end{aligned} \tag{4.41}$$

Now notice that both f_1 and f_5 are space-like separated from y; thus, according to Local Agency,

$$\begin{aligned}&p(f_1 = 1, f_5 = 1|x_{1-3} = 2, y = 1)\\ =&p(f_1 = 1, f_5 = 1|x_{1-3} = 2, y = 2) = 0.\end{aligned} \tag{4.42}$$

This implies that

$$p(f_1 = 1, f_5 = 1, b = +|x_{1-3} = 2, y = 2) = 0. \tag{4.43}$$

Using Eq. (4.32), we arrive at

$$p(f_1 = 1, f_5 = 1|x_{1-3} = 2, y = 2, b = +) = 0, \tag{4.44}$$

as claimed.

4.3.2 The Peres–Mermin example

4.3.2.1 The Peres–Mermin noncontextuality argument

We briefly recap the Peres–Mermin magic square [277, 343, 344] proof of (state-independent) contextuality. Consider the following square of quantum observables for a 4-dimensional qudit:

$$\begin{bmatrix} A & a & \alpha \\ B & b & \beta \\ C & c & \gamma \end{bmatrix} = \begin{bmatrix} \sigma_z \otimes \mathbf{1} & \mathbf{1} \otimes \sigma_z & \sigma_z \otimes \sigma_z \\ 1 \otimes \sigma_x & \sigma_x \otimes \mathbf{1} & \sigma_x \otimes \sigma_x \\ \sigma_z \otimes \sigma_x & \sigma_x \otimes \sigma_z & \sigma_y \otimes \sigma_y \end{bmatrix}. \tag{4.45}$$

Crucially, the triples of observables in a given column or row are all compatible with each other, and thus can be jointly measured. Thus the scenario consists of six possible

joint measurements, called measurement contexts—one for each row and one for each column—and each observable appears in exactly two contexts (namely in one row and in one column). Moreover, the product of all three observables in any given context is $\pm\mathbf{1}$:

$$\begin{aligned} Aa\alpha = \mathbf{1}, \quad Bb\beta = \mathbf{1}, \quad Cc\gamma = \mathbf{1}, \\ ABC = \mathbf{1}, \quad abc = \mathbf{1}, \quad \alpha\beta\gamma = -\mathbf{1}. \end{aligned} \tag{4.46}$$

By KSNC, the values assigned to mutually commuting observables must obey equalities satisfied by the observables themselves, and so the product of the values assigned to the three observables in each row or column must be ± 1. More specifically, denoting the assignment to observable O by $v_O \in \{-1, 1\}$, one must have

$$\begin{aligned} v_A v_a v_\alpha = 1, \quad v_B v_b v_\beta = 1, \quad v_C v_c v_\gamma = 1, \\ v_A v_B v_C = 1, \quad v_a v_b v_c = 1, \quad v_\alpha v_\beta v_\gamma = -1. \end{aligned} \tag{4.47}$$

But these six equations cannot be satisfied by any nine ± 1-valued assignments, since the first three equalities imply that

$$v_A v_a v_\alpha v_B v_b v_\beta v_C v_c v_\gamma = 1, \tag{4.48}$$

while the last three imply that

$$v_A v_a v_\alpha v_B v_b v_\beta v_C v_c v_\gamma = -1. \tag{4.49}$$

The above arguments also establish a fact that we will utilize later, which is that the following 6 probability distributions

$$p(v_A v_a v_\alpha = -1) = 0, \tag{4.50a}$$
$$p(v_B v_b v_\beta = -1) = 0, \tag{4.50b}$$
$$p(v_C v_c v_\gamma = -1) = 0, \tag{4.50c}$$
$$p(v_A v_B v_C = -1) = 0, \tag{4.50d}$$
$$p(v_a v_b v_c = -1) = 0, \tag{4.50e}$$
$$p(v_\alpha v_\beta v_\gamma = 1) = 0 \tag{4.50f}$$

cannot simultaneously be marginals of a joint probability distribution

$$p(v_A, v_a, v_\alpha, v_B, v_b, v_\beta, v_C, v_c, v_\gamma). \tag{4.51}$$

(Indeed, every KSNC argument can be cast as a marginal problem of this sort [42].)

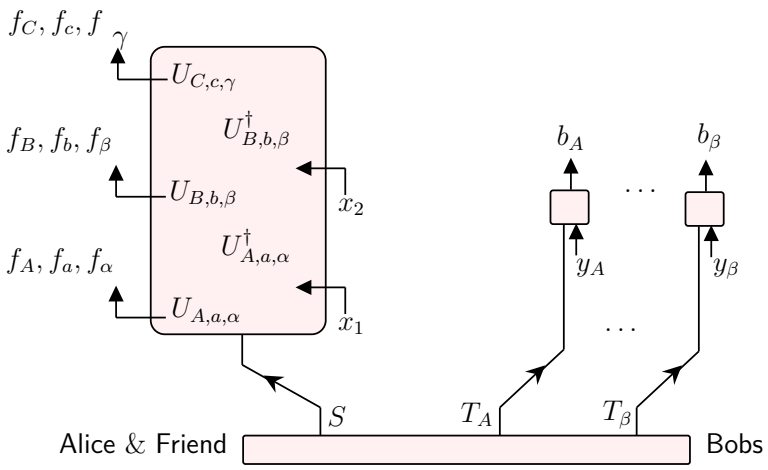

Figure 4.13: Schematic diagram of the LF Peres–Mermin scenario, see also Figure 4.14. All operations performed by Alice and Friend are space-like separated from all operations performed by the six Bobs; all operations performed by the six Bobs are further mutually space-like separated.

4.3.2.2 The LF Peres–Mermin protocol

We now leverage the Peres–Mermin KSNC argument to construct a novel scenario in which one can prove the LF no-go theorem.[8] This LF Peres–Mermin scenario involves 8 agents. The first two agents are Alice (denoted A) and her Friend (denoted F), of which Alice is the superobserver of the Friend. The remaining six agents are Bobs, denoted by B_A, B_a, B_α, B_B, B_b, B_β respectively. See Figure 4.14 for a cartoon representation of the set-up. In the following, we will describe the quantum protocol that will be used to prove the no-go theorem and use the assumption of Universality of Quantum Theory and the Born rule for empirical correlations; however, the scenario could also be described in a theory-independent way if one wished to derive robust inequalities for empirical data rather than simply a no-go theorem.

At the beginning of the experiment, F and the six Bobs share a 7-partite system ST where S is a qudit and T consists of 6 subsystems (denoted T_j for $j \in \mathbf{j} \coloneqq \{A, a, \alpha, B, b, \beta\}$), each of which is a qubit. The joint system ST is prepared in the

[8] Another extended Wigner's friend argument based on the Peres–Mermin KSNC argument was introduced in Ref. [18]; see also Ref. [20], where we argue that an additional assumption that we call Commutation Irrelevance is required there. Furthermore, unlike the LF Peres–Mermin protocol we develop here, the protocol in Ref. [18] cannot be used to prove a LF no-go theorem. The crucial difference is that our protocol includes an additional group of agents, namely, the Bobs, and allows the superobserver Alice to have measurement choices.

following entangled state:[9]

$$|\psi_0\rangle_{ST} \coloneqq \sum_{k_j=0,1,\forall j\in\mathbf{j}} \left(\prod_{j\in\mathbf{j}} \Pi_{j^{k_j}} |\psi\rangle\right)_S \bigotimes_{j\in\mathbf{j}} |k_j\rangle_{T_j}\,, \tag{4.54}$$

where $\Pi_{j^0} = \Pi_{j^\perp}$ is the projector into the subspace of the qudit S that yields outcome -1 for the observable $j \in \mathbf{j}$, while $\Pi_{j^1} = \Pi_j = \mathbf{1} - \Pi_{j^\perp}$, and $|\psi\rangle$ can be *any* 4-qubit quantum state.

System S is sent to the Friend while each T_j is sent to the corresponding B_j. On Alice's side, the Friend will measure the 9 observables $\{i\}_{i\in\mathbf{i}:=\{A,a,\alpha,B,b,\beta,C,c,\gamma\}}$ on system S. Specifically, she will first measure $\{A, a, \alpha\}$, then $\{B, b, \beta\}$, and finally $\{C, c, \gamma\}$. Whether she implements *each* triple of observables using a joint measurement (since they commute) or in a minimally disturbing sequential manner is not important for our proof of the LF no-go theorem. In either case, from AOE, the Friend will obtain an outcome, denoted f_i, for each observable $i \in \mathbf{i}$. (In the case where each triple of observables is measured using a single joint measurement, the f_is are the corresponding coarse-grained outcomes of the joint measurement.) We label the -1 outcome for the observable i by $f_i = -1$, and the other outcome by $f_i = 1$.

Under the assumption of Universality of Quantum Theory, we use $U_{A,a,\alpha}$, $U_{B,b,\beta}$, and $U_{C,c,\gamma}$ to denote the respective unitaries modeling the measurements for each triple of observables. The superobserver Alice is assumed to have perfect quantum control over the joint system of F and S. She will implement the inverse unitaries $U^\dagger_{A,a,\alpha}$ and $U^\dagger_{B,b,\beta}$ at specific points in time during the protocol.

The exact sequence of these quantum operations is shown in the schematic representation of the protocol in Fig. 4.13. Specifically, first, F measures A, a, and α on S, after which the superobserver Alice undoes these measurements by applying $U^\dagger_{A,a,\alpha}$.

[9] This initial entangled state, i.e., Eq. (4.54), can be constructed by

$$|\psi_0\rangle_{ST} = \left(\prod_{j=1}^{6} CNOT_{ST_j}\right) |\psi\rangle_S \bigotimes_{j=1}^{6} |0\rangle_{T_j} \tag{4.52}$$

where

$$CNOT_{ST_j} \coloneqq (\Pi_{j^\perp})_S \otimes \mathbf{1}_{T_j} + (\Pi_j)_S \otimes (|1\rangle\langle 0| + |0\rangle\langle 1|)_{T_j}. \tag{4.53}$$

Similar to Footnote 6, this entangled state can be constructed in a measure-prepare fashion around which the thought experiment could alternatively be constructed (although some extra complexity is introduced by the fact that since, as illustrated in Fig. 4.13, the B, b, β measurements are not done at the initial step of the protocol, and thus, more care needs to be taken regarding the space-time arrangement of the procedures).

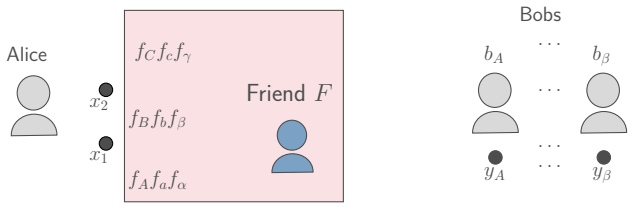

Figure 4.14: Cartoon representation of the LF Peres–Mermin scenario, see also Figure 4.13. Alice is a superobserver for her Friend, making choices x_1, x_2 whether to reveal the measurement outcomes obtained by the Friend, f_A, f_a, f_α and f_B, f_b, f_β, respectively. The spacelike separated Bobs make measurement choices $y_A, \ldots, y_\beta$ and obtain outcomes $b_A, \ldots, b_\beta$.

Then F measures B, b and β on S, followed by the undoing $U^\dagger_{B,b,\beta}$. Finally, F measures C, c, and γ on S.

In addition, the superobserver Alice has two binary choices x_1 and x_2. The x_1 choice is made right before she implements the inverse unitary $U^\dagger_{A,a,\alpha}$, and the x_2 choice is made right before she implements the inverse unitary $U^\dagger_{B,b,\beta}$. If $x_1 = 1$, Alice reveals measurement outcomes f_A, f_a and f_α to the world; if $x_1 = 2$, she simply continues the protocol. Similarly, when $x_2 = 1$, Alice reveals measurement outcomes f_B, f_b and f_β to the world; if $x_2 = 2$, she simply continues the protocol. If $x_1 = x_2 = 2$, the f_C, f_c and f_γ outcomes are revealed to the world at the end of the experiment.[10]

For each Bob B_j with any $j \in \mathbf{j}$, he has a binary measurement choice y_j. If $y_j = 1$, he measures system T_j in the computational basis; if $y_j = 2$, he measures system T in the $\pm$ basis, i.e.

$$\{|+\rangle\langle+|, |-\rangle\langle-|\}. \tag{4.55}$$

We denote B_j's respective absolute outcome by b_j.

We assume that all operations on Alice's side (including Alice and her Friend) are done at space-like separation from all operations by the Bobs; furthermore, all of the operations by the different Bobs are mutually space-like separated from each other.

Similar to the LF 5-cycle scenario, one can show that the Local Friendliness constraints on the LF Peres–Mermin scenario do not imply KSNC constraints on the compatibility graph associated with either the LF Peres–Mermin scenario or the KSNC Peres–Mermin scenario. Nevertheless, now we will show in Section 4.3.2 that the possibilistic proof of the failure of KSNC in the Peres–Mermin scenario can be mathematically turned into a proof of Local Friendliness no-go theorem.

[10] Similar to the LF 5-cycle case, any measurement outcome obtained after $x_n = 2$ is not used for the proof of the LF no-go theorem, so equivalently, whenever $x_n = 2$, Alice could also terminate the protocol on her side.

4.3.2.3 The no-go theorem

In this protocol, AOE demands the existence of the following conditional probability distribution:

$$P(f_{\mathbf{i}}, b_{\mathbf{j}}|x_1, x_2, y_{\mathbf{j}}) \tag{4.56}$$
$$:= p(f_A, f_a, f_\alpha, f_B, f_b, f_\beta, f_C, f_c, f_\gamma, b_A, b_a, b_\alpha, b_B, b_b, b_\beta|x_1, x_2, y_A, y_a, y_\alpha, y_B, y_b, y_\beta)$$

where $f_{\mathbf{i}}$, $b_{\mathbf{j}}$, and $y_{\mathbf{j}}$ are the respective short-hand notations for *all* elements in $\{f_i\}_{i\in\mathbf{i}}$, $\{b_j\}_{j\in\mathbf{j}}$, and that in $\{y_i\}_{j\in\mathbf{j}}$.

Below, we will consider the predictions made by quantum theory for various *empirical* correlations in the Peres–Mermin LF protocol, and we will then apply Local Agency to them to show that

$$p(f_A f_a f_\alpha = -1|x_1 = x_2 = 2, y_{\mathbf{j}} = 2, b_{\mathbf{j}} = +) = 0, \tag{4.57a}$$
$$p(f_B f_b f_\beta = -1|x_1 = x_2 = 2, y_{\mathbf{j}} = 2, b_{\mathbf{j}} = +) = 0, \tag{4.57b}$$
$$p(f_C f_c f_\gamma = -1|x_1 = x_2 = 2, y_{\mathbf{j}} = 2, b_{\mathbf{j}} = +) = 0, \tag{4.57c}$$
$$p(f_A f_B f_C = -1|x_1 = x_2 = 2, y_{\mathbf{j}} = 2, b_{\mathbf{j}} = +) = 0, \tag{4.57d}$$
$$p(f_a f_b f_c = -1|x_1 = x_2 = 2, y_{\mathbf{j}} = 2, b_{\mathbf{j}} = +) = 0, \tag{4.57e}$$
$$p(f_\alpha f_\beta f_\gamma = 1|x_1 = x_2 = 2, y_{\mathbf{j}} = 2, b_{\mathbf{j}} = +) = 0, \tag{4.57f}$$

where $y_{\mathbf{j}} = 2$ is the short-hand notation for $y_j = 2, \forall j \in \mathbf{j}$, and similarly for $b_{\mathbf{j}} = +$.

These six equations constitute constraints on six marginals of the joint distribution

$$p(f_{\mathbf{i}}|x_1 = x_2 = 2, y_{\mathbf{j}} = 2, b_{\mathbf{j}} = +), \tag{4.58}$$

which is well-defined due to the existence of Eq. (4.56) together with the empirical predictions given by the Born rule that

$$\wp(b_{\mathbf{j}} = +|x_1 = x_2 = 2, y_{\mathbf{j}} = 2) > 0. \tag{4.59}$$

However, from the Peres–Mermin KSNC proof in Section 4.3.2, we know that these six constraints, i.e., Eqs. (4.57a) to (4.57f), cannot be simultaneously satisfied (as they are analogous to Eqs (4.50)). Therefore, there can be no joint distribution of the form in Eq. (4.58) that is consistent with the empirical predictions given by the Born rule and Local Agency. Hence the no-go theorem.

It remains only to establish Eqs. (4.57a) to (4.57f). The steps for proving Eqs. (4.57a) to (4.57c) are similar to those for proving Eqs. (4.29a) to (4.29d) in Section 4.3.1, and

the steps for proving Eqs. (4.57d) and (4.57f) are similar to those for proving Eq. (4.30) in Section 4.3.1. A proof can be found in Appendix I of Ref. [21].

4.3.3 Translating any possibilistic KSNC argument into a LF argument

We have now seen two examples of how possibilistic proofs of the failure of KSNC can be translated into LF arguments. More generally, one has the following:

Theorem 4. *Every quantum possibilistic proof of the failure of Kochen–Specker noncontextuality can be mathematically translated into a proof of the failure of Local Friendliness in a corresponding extended Wigner's friend scenario.*

We prove this in Appendix J of Ref. [21], using a construction analogous to the above constructions (where all Bobs only need to perform either the computational basis measurement or the complementary $\pm$ measurement).

Note that in these LF constructions, one introduces extra measurements beyond those in the contextuality scenarios[11]; in particular, the measurements made by Bobs on qubits T_j. However, one can avoid the addition of these additional measurements (and any choices of Alice or of the Bobs) if one appeals to the assumption of *Commutation Irrelevance* of Section 4.1 rather than Local Agency. For details, see Appendix K of Ref. [21].

One could alternatively provide a different sort of translation from KSNC arguments to LF arguments by first translating the KS construction into a Bell scenario using known procedures [346–349], and then using the mapping from Bell scenarios to LF scenarios given by reversing the translation given in Section 4.2.7. (Note that the translation from a KSNC argument to a Bell argument also involves adding additional measurements in the case where the initial KSNC argument is state-dependent [346].)

4.4 An extended Wigner's friend paradox based on generalized contextuality

The Local Friendliness (LF) no-go theorem [12–14] shows a contradiction between the predictions of quantum theory and the assumptions of Absoluteness of Observed Events

[11] It is commonplace for translations between different scenarios to involve slightly different states and measurements; see for example Ref. [252, 345, 346] for other examples.

– ascribing single and objective values to observed measurement outcomes – and Local Agency – a strengthened version of no-signaling that also applies to situations involving outcome correlations that cannot be verified by a single agent. As discussed in Chapter 2, the LF no-go theorem can be conceptualized as combining a Wigner's friend experiment with a Bell scenario and, assuming that the LF scenario can be realized, results in an even stronger theorem than Bell's theorem, given that the LF assumptions are weaker than the ones of Bell's theorem [265]. One way to ensure that the LF scenario can be realized is to assume the Universality of Quantum Theory, in that superobservers exist. The sense in which the LF no-go theorem is said to be stronger than Bell's theorem is exemplified in Ref. [17], where the authors show that, in standard LF scenarios, the polytope of correlations consistent with Local Friendliness (the conjunction of Absoluteness of Observed Events and Local Agency) is strictly larger than the polytope of locally causal correlations.

In this section, we combine Wigner's friend experiments with scenarios manifesting generalized contextuality [43], one of the leading notions of nonclassicality. Specifically, we consider the simplest scenario showing generalized contextuality [260], which is a prepare-and-measure scenario on a single system involving four preparations and two measurements, and we include quantum observers. This theorem demonstrates the inconsistency of quantum theory with the joint assumptions of Absoluteness of Observed Events and *Noncontextual Agency*, the latter being a weaker version of noncontextuality, in the same way that Local Agency is a weaker version of local causality. Our result generalizes the Local Friendliness no-go theorem and, granted that the scenario can be realized, is stronger than no-go theorems based on generalized noncontextuality. In designing the experiment we leverage a known correspondence between proofs of nonlocality in Bell scenarios and proofs of contextuality in prepare-and-measure scenarios [251, 281, 350–352]. In the same way in which, under the assumption that the scenario can be realized, the LF no-go theorem is stronger than Bell's theorem, we obtain a stronger no-go theorem than the one associated with generalized contextuality.

This section contains several copied parts of Ref. [22].

4.4.1 Bell nonlocality and generalized contextuality

In a Bell scenario, two parties, Alice and Bob, share a bipartite system and each perform a measurement on their respective parts. Alice's and Bob's measurement choices are denoted by x and y, respectively, with outcomes a and b. Repeating the protocol many

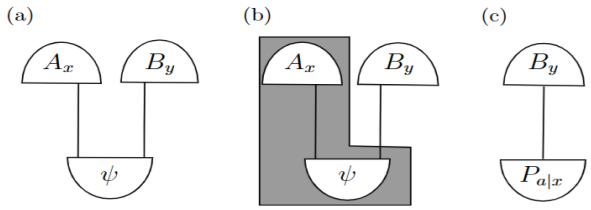

Figure 4.15: (a) Bell scenario; (b) Bell scenario as a prepare-and-measure scenario; (c) prepare-and-measure scenario. Subscripts denote the preparation and measurement variables, and ψ the bipartite state.

times allows for the collection of measurement statistics $\wp(a, b|x, y)$.[12] These correlations are consistent with the existence of an ontological model satisfying *local causality* [35] if they can be expressed as follows:

$$\wp(a, b|x, y) = \sum_{\lambda} p(\lambda)p(a|x, \lambda)p(b|y, \lambda), \tag{4.60}$$

where λ represents the ontic state in the ontological model. An equivalent way to require that Eq. (4.60) is satisfied is to demand the existence of a joint distribution $p(a_1, a_2, b_1, b_2)$, where $a_x = a|x, b_y = b|y$, for all possible measurement outcomes and settings, from which the observed probabilities $\wp(a, b|x, y)$ can be obtained by marginalizing over the unperformed measurements [42, 342], for example, $\wp(a, b|x = 2, y = 2) = \sum_{a_1,b_1} p(a_1, a_2, b_1, b_2)$. As we have encountered several times in the previous sections and chapters, Bell's theorem famously proves that quantum theory manifests *nonlocality*, *i.e.*, it does not admit of a locally causal ontological model [35].

A Bell scenario can also be conceptualized as a prepare-and-measure scenario, where the preparation stage on Bob's side is determined by steering from Alice's setting and outcome (x, a), and the measurement stage consists of Bob's measurement y and outcome b [281, 350–352] (see Fig. 4.15). In this setting, the nonlocality of quantum theory turns into the inconsistency of quantum theory with a generalized noncontextual ontological model [43].

As introduced in Section 2.2, generalized noncontextuality requires that operational equivalences predicted by a physical theory are preserved in the ontological model of the theory. In the context of local causality, no-signaling is the operational equivalence that is preserved in the ontological model. Specifically, with respect to a in a Bell scenario, no-

[12] We here introduce, similarly as in the previous section, a notational difference between empirical probability distributions that can be obtained by a single agent from the collected data, denoted with $\wp$, and theoretical probability distributions required to exist given certain assumptions, but potentially not observable by any agent, denoted with p.

signaling implies $\wp(a|x,y) = \wp(a|x)$ and its preservation in the ontological model reads as $p(a|\lambda,x,y) = p(a|\lambda,x)$. This assumption, known as parameter independence, combined with outcome independence – *e.g.*, with respect to a, $p(a|\lambda,b,x,y) = p(a|\lambda,x,y)$ – defines local causality[13].

The simplest proof that quantum theory manifests *contextuality*, *i.e.*, does not admit of a generalized noncontextual ontological model, can be constructed in the *simplest nontrivial scenario* [260–262]. This is a prepare-and-measure scenario consisting of two tomographically complete measurements, *e.g.*, the Pauli X and Z measurements on a qubit, and four preparations, *e.g.*, $P_\theta = \cos(\theta/2)\,|0\rangle + \sin(\theta/2)\,|1\rangle$, for $\theta = \pi/4, 3\pi/4, 5\pi/4, 7\pi/4$. These are represented on the Bloch disk in Fig. 4.16. The preparations satisfy the following operational equivalence,

$$\frac{1}{2}\wp(k|M,P_{\frac{\pi}{4}}) + \frac{1}{2}\wp(k|M,P_{\frac{5\pi}{4}}) = \frac{1}{2}\wp(k|M,P_{\frac{3\pi}{4}}) + \frac{1}{2}\wp(k|M,P_{\frac{7\pi}{4}}).$$

However, there does not exist an ontological model where these also provide an ontological equivalence.

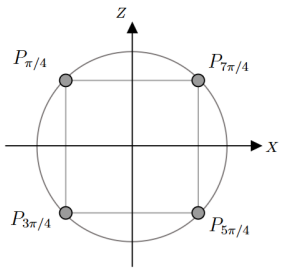

Figure 4.16: The simplest nontrivial scenario [260], involving four preparations (gray dots) and two measurements (the two axes) that provide a maximal violation of the noncontextuality inequalities.

A strategy obtaining a maximum quantum violation in Bell scenario can be mapped to the simplest scenario with the specific choices of preparations and measurements above. In the former case, one shows nonlocality; in the latter, preparation contextuality.

4.4.2 Rephrasing the Local Agency assumption

Recall the Local Friendliness scenario of Section 2.5, illustrated in Fig. 4.17, involving Charlie and Debbie and their superobservers Alice and Bob who make choices x, y.

Given the setup, let us rephrase the assumption of Local Agency.

[13] Note that the assumption of measurement independence, $p(\lambda|x,y) = p(\lambda)$, which is also required to prove Bell's theorem, is already built into the definition of ontological models [228].

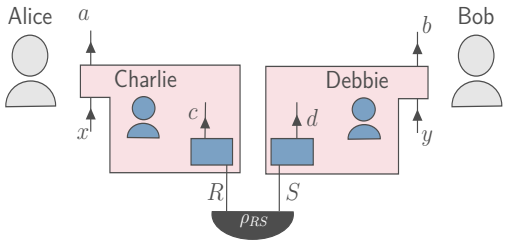

Figure 4.17: Local Friendliness setup [27].

Assumption 7 (Local Agency)**.** No-signaling outside the future light cone, which would be verified by a hypothetical agent with access to all the relevant variables, still holds even if it cannot be verified by a single agent.

The above formulation of Local Agency is equivalent to the standard one [13]:

> "The only relevant events correlated with an intervention are in its future light cone."

Indeed, the choice settings ("the interventions") can only be correlated with the relevant events that lie within their future light cone, *even if all the events cannot be observed by a single agent*. In this sense, Local Agency can be read as a stronger version of standard no-signaling. In the LF scenario, Local Agency implies, for example, that $p(c, b|x, y = 2) = p(c, b|y = 2)$, as none of c, b are obtained in the future of the choice x. Notice how this is different from requiring the disjoint conditions to hold: $\wp(b|x, y = 2) = \wp(b|y = 2)$ – an instance of no-signaling – and $\wp(c|x, y) = \wp(c)$ – an instance of no-signaling to the past, or no-superdeterminism. Local Agency is stronger than standard no-signaling and no-superdeterminism insofar as it requires the joint independence, which, in addition to the no-signaling and no-superdeterminism above, also requires $p(b|c, x, y = 2) = p(b|c, y = 2)$.

Local Agency can be motivated by a commitment to relativistic principles and the belief that these should hold even if they cannot be directly verified by a single agent. It can also be justified as an instance of no operational fine-tuning [228] in a universe where AOE holds: if events are absolute and a theory prescribes no-signaling, the latter should also hold even if it cannot be verified by a single agent. Put differently, no-signaling should hold even if its verification were possible only for a hypothetical agent who could access all relevant variables by cheating the protocol. Let us call this agent Eve and imagine that she has access to outcomes b, c and choices x for $y = 2$. Beyond the standard no-signaling instances $\wp(b|x, y = 2) = \wp(b|y = 2)$ and $\wp(c|x, y) = \wp(c)$, Eve can now verify an additional no-signaling instance: $p(b, c|x = 2, y = 2) = p(b, c|x = 1, y = 2)$,

i.e., if Eve were to access only $b, c, y = 2$, she would not get any information on x. Local Agency demands that such no-signaling conditions verified by Eve must remain valid even in her absence, and even if the relevant events are jointly inaccessible in the sense that they are not jointly observed by any agent, yet still exist by virtue of AOE.

It is worth noting that while Local Agency is stronger than no-signaling, it is weaker than local causality, as shown in Ref. [13]. This is further evident in scenarios where the polytope of correlations implied by Local Friendliness (the conjunction of AOE and Local Agency) is strictly larger than the polytope of locally causal correlations, as demonstrated in Ref. [17].

4.4.3 Noncontextual Friendliness no-go theorem

The experimental setup, illustrated in Fig. 4.18, is described as follows. Alice prepares a qubit S in the state P_a, for values of $a = 0, 1$ each happening with probability $1/2$. She sends the prepared system to her friend Charlie, who performs a binary measurement C on S, where the two possible outcomes are labeled by 0 and 1. The measurement is modeled unitarily as U_C. After Charlie has performed his measurement, depending on the value of the variable $x \in \{1, 2\}$, Alice makes one of two choices. If $x = 1$ she takes Charlie's outcome to be her outcome. If $x = 2$ she acts as a superobserver and undoes Charlie's measurement by applying $U_C^\dagger$, thereby erasing the outcome record of c. Notice that, in this latter case, unlike the LF setup, she does not perform a measurement A on S, after erasing the outcome of Charlie. Subsequently, the system is passed to Debbie who performs a measurement D on the system. This measurement is modeled unitarily as U_D. After Debbie has performed her measurement, depending on the value of the variable $y \in \{1, 2\}$, Bob makes one of two choices. If $y = 1$ he takes Debbie's outcome to be his outcome. If $y = 2$ he acts as a superobserver and undoes Debbie's measurement by applying $U_D^\dagger$, thereby erasing the outcome record of d, and subsequently performs a measurement denoted with B on S yielding outcome b.

Given this setup, let us state our assumption of Noncontextual Agency, replacing Local Agency. Accordingly, we name the conjunction of AOE and Noncontextual Agency *Noncontextual Friendliness* (NF).

Assumption 8 (Noncontextual Agency)**.** Any operational equivalence, which would be verified by a hypothetical agent with access to all the relevant variables, still holds even if it cannot be verified by a single agent.

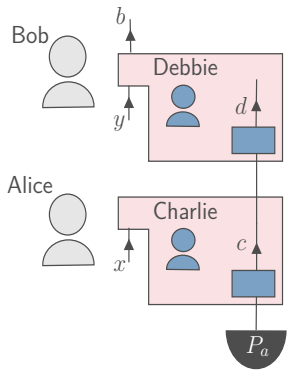

Figure 4.18: Noncontextual Friendliness setup.

Notice the similarity between Noncontextual Agency and Local Agency of Assumption 7: the former can be viewed as a special case of the latter, where the operational equivalence considered is specifically no-signaling. Just as Local Agency is a stronger version of no-signaling, Noncontextual Agency is a stronger version of standard operational equivalence.[14] In the context of the setup considered here, standard operational equivalence pertains to setting choices. More precisely, they are about the preparations associated with choices $x = 1$ and $x = 2$: $\wp(d|x, y = 1) = \wp(d|y = 1)$ and $\wp(b|x, y = 2) = \wp(b|y = 2)$. Noncontextual Agency additionally implies that these equalities also hold when conditioned on c, *i.e.*, $p(d|c, x, y = 1) = p(d|c, y = 1)$ and $p(b|c, x, y = 2) = p(b|c, y = 2)$, despite the fact that no single observer can verify them, since when $x = 2$, the outcome c is erased. Thus, similarly to Local Agency, Noncontextual Agency lifts the disjoint conditional independences of $d|y$ and $c|y$ on x to a joint conditional independence of $c, d|y$ on x (and similarly for $b, c|x, y$).

Regarding the name "Noncontextual Agency", the term noncontextual is used similarly to generalized noncontextuality: just as generalized noncontextuality requires that operational equivalences are preserved as ontological identities within an ontological model explaining the physical theory, here noncontextual indicates that operational equivalences are preserved at the level of the Absoluteness of Observed Events – and thus remain valid even when they are operationally inaccessible, in the sense that no single agent can verify them. In other words, it represents a form of noncontextuality formulated not within the framework of ontological models [259], but under the sole assumption that the underlying reality satisfies AOE. This terminology parallels that of Local Agency,

[14] As is the case for Noncontextuality, Noncontextual Agency relies on operational equivalences that are defined once the theory – and hence the relevant systems and the set of in principle distinguishing measurements – has been specified; it is therefore subject to the "system loophole" formulated in [353].

where local is inspired by the notion of local causality in Bell's theorem (specifically, its parameter independence aspect, sometimes referred to as locality [354]), but where no-signaling is required to hold only at the AOE level – *i.e.*, in a world where observed events are absolute – rather than in an ontological model.

Noncontextual Agency can be seen as an instance of no operational fine-tuning [228] in a universe where AOE holds. Indeed, as with Local Agency, one can motivate it by arguing that if a theory prescribes an operational equivalence, such an equivalence should also hold even if it cannot be verified by a single agent. Put differently, an operational equivalence should hold even if its verification were possible only for a hypothetical agent who could access all relevant variables by cheating the protocol.

Let us call this agent Eve, and imagine that she has access to outcomes c, d and choices x, y. Since c lies in the past of x, y, Eve can verify, in accordance with the quantum predictions, that $\wp(c|x, y) = \wp(c)$; in fact, this independence can also be verified without referring to Eve, as shown in Section 4.2.2 of Ref. [33]. If, moreover, Eve cheats the protocol by storing Charlie's outcome c in a device before any potential undoing action is performed on Charlie's lab, she can then verify that $p(d|c, x, y) = p(d|c, y)$. Indeed, if Eve has access to c and d, the probabilities $p(d|c, x, y)$ correspond to preparing system S (the system Charlie measured) in the post-measurement state P_c associated with outcome c, followed by Alice's transformation, which depends on x, and Debbie's measurement. For $x = 1$, Alice takes Charlie's outcome to be hers thus not affecting the system S, and for $x = 2$ she applies $U_C^\dagger$, erasing c. However, as Eve has already copied c into her memory, which remains unaffected by Alice's operations, the undoing $U_C^\dagger$ does not alter P_c. Thus, in both cases $x = 1$ and $x = 2$, Eve observes that P_c is unaltered until Debbie applies her measurement, confirming that $p(d|c, x, y) = p(d|c, y)$. A similar argument establishes that $p(b|c, x, y) = p(b|c, y)$. Noncontextual Agency demands that the operational equivalences verified by Eve must remain valid even in her absence, and even if the relevant events are jointly inaccessible in the sense that they are not jointly observed by any agent, yet still exist by virtue of AOE.

It is worth noting that while Noncontextual Agency is stronger than operational equivalence, it is weaker than generalized noncontextuality: as shown in Appendix B of Ref. [22], one can construct a scenario showing that the polytope of NF correlations is strictly larger than the polytope of noncontextual correlations, mirroring the proof strategy of Ref. [17].

We can now state the NF no-go theorem.

Theorem 5 (Noncontextual Friendliness no-go theorem)**.** *If a superobserver can perform arbitrary quantum operations on an observer and its environment, then no physical theory can satisfy Noncontextual Friendliness.*

Proof. Let us consider the NF scenario of Fig. 4.18 with the following specifications (the preparations and measurements employed will be the same as in the simplest nontrivial scenario). Alice chooses preparations $P_0 = P_{\pi/4}$ and $P_1 = P_{5\pi/4}$, for $a = 0, 1$, respectively. Charlie measures in the basis given by $P_{3\pi/4}$ and $P_{7\pi/4}$. Debbie's measurement D is the Pauli Z measurement. When $y = 2$, Bob undoes Debbie's measurement and performs a Pauli X measurement on S. At the end of the experiment the following empirical correlations of observed outcomes are obtained,

$$\begin{aligned} &\wp(c, d \mid x = 1, y = 1), \\ &\wp(c, b \mid x = 1, y = 2), \\ &\wp(a, d \mid x = 2, y = 1), \\ &\wp(a, b \mid x = 2, y = 2). \end{aligned} \tag{4.61}$$

By AOE and Noncontextual Agency, we can rewrite these empirical correlations as follows,

$$\wp(c, d \mid x = 1, y = 1) = p(c, d|x = 2, y = 2), \tag{4.62}$$
$$\wp(c, b \mid x = 1, y = 2) = p(c, b|x = 2, y = 2), \tag{4.63}$$
$$\wp(a, d \mid x = 2, y = 1) = p(a, d|x = 2, y = 2), \tag{4.64}$$
$$\wp(a, b \mid x = 2, y = 2) = p(a, b|x = 2, y = 2). \tag{4.65}$$

AOE is used to consider joint probability distributions when c or d are involved and possibly erased. To obtain Eq. (4.62), we first appeal to an instance of Noncontextual Agency that enforces no-superdeterminism, $p(c, d|x, y){=}p(c, d|x)$, yielding $p(c, d|x{=}1, y{=}1){=}p(c, d|x{=}1, y{=}2)$; we can then apply the rule of conditional probability and consider two other instances of Noncontextual Agency (the first again enforcing no-superdeterminism): $p(c|x, y){=}p(c)$ and $p(d|c, x, y){=}p(d|c, y)$, thus obtaining

$$\begin{aligned} &\wp(c, d|x{=}1, y{=}1) = p(c|x{=}1, y{=}1)p(d|c, x{=}1, y{=}1) \\ &= p(c|x{=}2, y{=}2)p(d|c, x{=}2, y{=}2) = p(c, d|x{=}2, y{=}2). \end{aligned}$$

Similar instances of Noncontextual Agency, but with variables c, b – namely $p(c|x, y) = p(c)$ and $p(b|c, x, y) = p(b|c, y)$ – lead to Eq. (4.63),

$$\begin{aligned}&\wp(c, b|x{=}1, y{=}2) = p(c|x{=}1, y{=}2)p(b|c, x{=}1, y{=}2)\\&= p(c|x{=}2, y{=}2)p(b|c, x{=}2, y{=}2) = p(c, b|x{=}2, y{=}2).\end{aligned}$$

Eq. (4.64) derives from an instance of Noncontextual Agency enforcing no-superdeterminism, $p(a, d|x, y) = p(a, d|x)$.

By assuming AOE, we can assert that all the correlations in Eqs. (4.62),(4.63),(4.64) and (4.65) can be obtained from a single joint probability distribution $p(a, b, c, d|x = 2, y = 2)$ via marginalization. However, recall that the preparations and measurements used to obtain these correlations correspond to those in the simplest scenario that achieve the maximum violation of noncontextuality inequalities [260]. Through the mapping to the Bell scenario, they are also equivalent to those achieving the maximum violation of Bell inequalities. By Fine's theorem [42, 247], such a joint probability distribution cannot exist. Let us stress that the NF assumptions are weaker than those used in the noncontextuality no-go theorem proven in the simplest scenario [260]. Only the empirical correlations in the scenarios are the same. Notice how the present proof mirrors the proof of the LF no-go theorem, with Noncontextual Agency replacing Local Agency.

4.4.4 Discussion

In this section, we have shown that the Local Friendliness no-go theorem can be generalized to the Noncontextual Friendliness no-go theorem. Just as no-signaling is a specific instance of an operational equivalence, Local Agency is a specific instance of Noncontextual Agency. Furthermore, just as the LF no-go theorem is stronger than Bell's theorem, granted that the LF scenario can be realized, the NF no-go theorem is stronger than theorems based on generalized noncontextuality, particularly the one leveraging the simplest scenario, granted that the NF scenario can be realized. Local Agency strengthens no-signaling by requiring it to hold even when it cannot be verified by a single agent. Similarly, Noncontextual Agency extends operational equivalence to cases where no single agent can verify the equivalence.

Key to establishing these connections is the rephrasing of Local Agency, highlighting how it can be understood as a stronger version of no-signaling. Specifically, Local Agency can be interpreted as no-signaling verified by a hypothetical agent who is capable of breaking the protocol, accessing all relevant variables, and confirming the condition.

Recognizing that no-signaling is itself an instance of operational equivalence naturally leads to the mirrored definition of Noncontextual Agency. The assumption of Noncontextual Agency encompasses the newly introduced assumption of Commutation Irrelevance of Section 4.1.

It is worth discussing here the conceptual significance of the result of this section, which primarily arises from the conceptual status of Noncontextual Agency. In our formulation, the assumption of Noncontextual Agency constitutes an instance of the methodological principle of no fine-tuning: in a world where AOE holds, it would be conspiratorial for operational equivalences to apply only when verified by a single agent, but to fail otherwise. It is indeed natural (i.e., non-conspiratorial) to assume that, in a world satisfying AOE, the theory continues to predict equivalences to hold for correlations involving jointly inaccessible outcomes. Moreover, as we emphasize through the introduction of the hypothetical agent, if one were hypothetically able to access all relevant variables simultaneously, one would indeed verify the operational equivalence (albeit in a different protocol that breaks the one under consideration).

Notice that the above considerations also apply to the case of Local Agency, which is an instance of Noncontextual Agency. Consequently, Local Agency shifts from being justified by a metaphysical assumption about the nature of the underlying causal structure—namely, that it must be relativistic—to being an instance of the methodological principle of no fine-tuning. As such, it carries a lighter ontological weight. In turn, the lesson drawn from the LF and NF no-go theorems becomes clear: if one believes that the LF and NF scenarios can indeed be realized – for instance, by endorsing the universality of quantum theory[15] – and accepts the no fine-tuning (or no-conspiracy) principle – which, following many others [228, 229, 355], we regard as methodologically indispensable – then one is compelled to conclude that observed events are relative; that is, one must abandon AOE, the sole remaining metaphysical assumption in the theorem.

Interpreting Local Agency as an instance of Noncontextual Agency also allows us to dig deeper into the justification of Local Agency itself. Even the justification based on the belief in an underlying relativistic causal structure—one that is assumed to hold also for inaccessible outcomes—can only be motivated by an appeal to the no fine-tuning principle in a world where AOE holds. Otherwise, there would be no reason to endorse it.

[15] In support of the universality of quantum theory, we note that there is currently no experimental evidence against it, nor any indication of regimes in which it might fail. Moreover, even if quantum theory were not universal, one would still be required to specify precisely the domains of its applicability and to provide a clear account of how the transition to alternative physical descriptions occurs in other regimes.

Why should one believe that no-signaling also holds for jointly inaccessible outcomes, and that the relativistic causal structure extends to them? Because, in a world where AOE holds, it would otherwise be conspiratorial not to.

As far as we are aware, the only work that explicitly discusses no fine-tuning in the context of an extended Wigner's friend scenario is Ref. [33], although the authors do so within the framework of causal models. They show that, even when one allows for a nonclassical causal model and assumes AOE, it remains impossible to account for the statistics of the LF scenario without invoking causal fine-tuning [223, 226] – a result which bears a close relation to ours. We leave a detailed investigation of this connection for future work.

Another interesting avenue would be to explore proofs of the NF no-go theorem in scenarios beyond the one considered here, which is based on the simplest scenario. More broadly, it would be valuable to generalize the NF no-go theorem by considering examples of fine-tunings that go beyond those linked to contextuality [228], such as violations of bounded ontological distinctness [262, 356]. Additionally, it would be worthwhile to investigate NF scenarios that are realizable within post-quantum theories – for example those proposed in [23, 25, 357].

— 5 —

Are free choices absolute, when Pusey–Barrett–Rudolph are friends of Wigner?

The Local Friendliness no-go theorem, which we have encountered several times in the previous chapters, can be read as an argument against the absolute outcomes of performed measurements, under a locality and universality assumption. Proofs of this no-go theorem relied on correlations violating Bell locality, and, as we have shown in Chapter 4, proofs can be found based on noncontextuality as well. In this chapter, we provide an entirely different extended Wigner's friend scenario based on the Pusey–Barrett–Rudolph scenario, which tackled questions on the reality of the quantum state. This scenario can be interpreted as an argument against the absoluteness of free choices, under the same locality assumption as in the Local Friendliness no-go theorem.

5.1 The Pusey–Barrett–Rudolph argument

The Pusey–Barrett–Rudolph theorem [38] is a result against an epistemic reading of the quantum state in the *ontological models framework*. As reviewed in Section 2.2.2, in the framework of ontological models, a physical system is associated with one or more ontic states λ about which a measurement can reveal certain information. It may occur that full knowledge of a preparation P of the system does not specify a unique ontic state λ, but only a distribution $\mu_P(\lambda)$ over such ontic states, referred to as the ontic distribution. A measurement M is represented by a probability distribution $p(\xi|\lambda, M)$

over its outcomes ξ and the empirical probability of obtaining ξ upon a measurement M given preparation P is given by $p(\xi|M,P) = \sum_\lambda p(\xi|M,\lambda)\mu_P(\lambda)$.

The question of whether the quantum state is real or not can be interpreted as follows within this framework. The quantum state is said to be real, or *ontic*, if each pair of distinct, pure quantum states $|\psi_0\rangle, |\psi_1\rangle$ have non-overlapping distributions, $\mu_{\psi_0}(\lambda)\mu_{\psi_1}(\lambda) = 0$. The quantum state is said to be *epistemic* if it is not ontic, i.e. if there exists a pair of operationally distinguishable quantum states $|\psi_0\rangle, |\psi_1\rangle$ that have overlapping ontic distributions, $\mu_{\psi_0}(\lambda)\mu_{\psi_1}(\lambda) > 0$. A quantum state being epistemic captures the idea that it represents mere knowledge; if the ontic distributions $\mu_{\psi_0}(\lambda), \mu_{\psi_1}(\lambda)$ corresponding to preparations of two different pure quantum states $|\psi_0\rangle, |\psi_1\rangle$ have overlapping support, then these quantum states cannot refer to a physical property of the physical system.

The Pusey–Barrett–Rudolph argument against epistemic states in this framework goes as follows, following Ref. [38] closely. Consider two copies of a preparation device that operate independently and can each prepare a quantum system in the state $|0\rangle$ or $|+\rangle$. Here $|0\rangle, |1\rangle$ is a basis of a qubit, $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and for later convenience we also define $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$. The possible preparations arising from the two devices jointly are thus $|0\rangle\otimes|0\rangle, |0\rangle\otimes|+\rangle, |+\rangle\otimes|0\rangle$ and $|+\rangle\otimes|+\rangle$.

Beyond the ontological models framework assumption (i), it is assumed that (ii) a form of *preparation independence* holds [231, 358], i.e., that the joint ontic distribution of multipartite preparations leading to product quantum states factorizes, $\mu_{P_A P_B}(\lambda_A, \lambda_B) = \mu_{P_A}(\lambda_A)\mu_{P_B}(\lambda_B)$. Note that this property is only required for product states, and need not be assumed for entangled states.

Next, an entangled measurement is performed upon the two systems, which projects onto the four orthogonal states

$$\begin{aligned}
|\xi_1\rangle &= \frac{1}{\sqrt{2}}(|0\rangle\otimes|1\rangle + |1\rangle\otimes|0\rangle),\\
|\xi_2\rangle &= \frac{1}{\sqrt{2}}(|0\rangle\otimes|-\rangle + |1\rangle\otimes|+\rangle),\\
|\xi_3\rangle &= \frac{1}{\sqrt{2}}(|+\rangle\otimes|1\rangle + |-\rangle\otimes|0\rangle),\\
|\xi_4\rangle &= \frac{1}{\sqrt{2}}(|+\rangle\otimes|-\rangle + |-\rangle\otimes|+\rangle).
\end{aligned} \tag{5.1}$$

Now if the ontic distributions of the quantum states $|0\rangle, |+\rangle$ overlap, then there is a nonzero probability $q^2 > 0$ that an ontic state λ is prepared that is compatible with all possible preparations listed above. Each of these states $|\xi_i\rangle$ is orthogonal to a different

one of the given preparations, implying that the corresponding measurement outcome ξ_i has probability zero of occurring when a prepared state is orthogonal to $|\xi_i\rangle$. Namely, $|\xi_1\rangle$ is orthogonal to $|0\rangle \otimes |0\rangle$, $|\xi_2\rangle$ is orthogonal to $|0\rangle \otimes |+\rangle$, $|\xi_3\rangle$ to $|+\rangle \otimes |0\rangle$ and $|\xi_4\rangle$ to $|+\rangle \otimes |+\rangle$. This fact immediately gives the contradiction; when given such λ compatible with all possible preparations, the measurement device is unsure which preparation had been performed, and thus runs the risk of giving a measurement outcome for which the actual preparation had probability zero of occurring.

This shows that the quantum states $|0\rangle, |+\rangle$ must have zero overlap in their ontic distributions. A generalization of this argument shows that in fact any pair of distinguishable quantum states must have zero overlap [38], proving that the quantum state must be ontic, modulo the assumptions $(i) - (ii)$ presented above. Arguments that loosen the preparation independence assumption (ii) can be found in Refs. [231, 358, 359].

5.2 Are free choices absolute?

The scenario that we will consider, sketched in Fig. 5.1, combines Wigner's friend with the Pusey–Barrett–Rudolph argument to derive a contradiction with the Local Friendliness assumptions, now applied to free choices.

Two friends, Alice and Bob, residing in sealed labs and modeled as quantum systems by their superobserver Wigner, each make a 'free choice' $a, b \in \{0, 1\}$. For now, we assume that Alice and Bob choose 0 and 1 with equal probabilities $1/2$.[1] They next encode their choices a, b in the computational bases of qubits T_A, T_B that they send to Charlie and Debbie, respectively.[2]

We can model the state at this point by

$$|\Phi\rangle = |\phi_0\rangle_{T_A L_A} \otimes |\phi_0\rangle_{T_B L_B}, \tag{5.2}$$

with, for $I = A, B$,[3]

$$|\phi_0\rangle_{T_I L_I} = \frac{1}{\sqrt{2}} \Big(|00\rangle_{T_I L_I} + |11\rangle_{T_I L_I} \Big). \tag{5.3}$$

Here $|0\rangle_{L_A}$ and $|1\rangle_{L_A}$ denote the states of Alice's lab when she chose $a = 0$ and $a = 1$, respectively, and similarly for Bob.

[1] In fact, more precisely, we assume for now that the amplitudes of their choices are given by $1/\sqrt{2}$.

[2] The systems T_A, T_B are the only systems that can escape Alice's and Bob's otherwise sealed labs.

[3] Here and later, for ease of notation, we omit the tensor product, i.e. we write $|00\rangle$ or $|0\rangle\,|0\rangle$ for $|0\rangle \otimes |0\rangle$.

After this, Wigner chooses $z = 1, 2$, where for $z = 1$ he reveals Alice's and Bob's choices (for example by measuring their labs in the computational basis) and for $z = 2$ he performs the entangled PBR measurement on Alice's and Bob's labs L_A, L_B, namely measuring $L_A \otimes L_B$ in the basis $|\xi_1\rangle$, $|\xi_2\rangle$, $|\xi_3\rangle$, $|\xi_4\rangle$. Wigner's outcomes are denoted by $w \in \{\xi_1, \xi_2, \xi_3, \xi_4\}$.

Having received the qubits T_A, T_B from Alice and Bob, Charlie and Debbie choose measurement settings x and y, respectively. For $x = 1$, Charlie decodes Alice's choice by measuring T_A in the computational basis, obtaining outcome $c \in \{0, 1\}$. For $x = 2$, Charlie measures T_A in the X-basis, obtaining outcome $c \in \{+, -\}$, and similarly for Debbie choosing $y = 1, 2$ to measure T_B in the Z– or X–basis. Charlie and Debbie do so spacelike separated from each other and from Wigner.

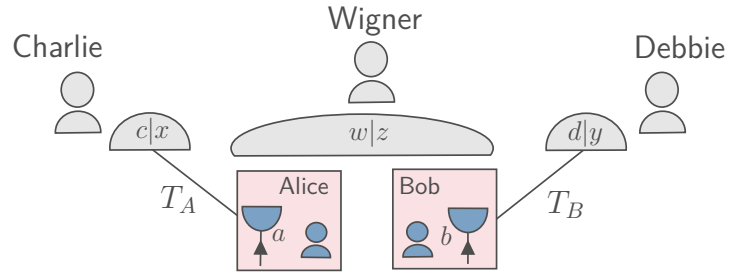

Figure 5.1: Protocol that combines Wigner's friend and the PBR argument to violate Local Friendliness. Two friends Alice and Bob each make a choice $a, b \in \{0, 1\}$. They encode their choices in the computational basis of qubits that they send to Charlie and Debbie, who choose x, y whether to decode their outcomes or measure their received qubit in the X-basis. Wigner is a superobserver for Alice and Bob, and chooses $z = 1, 2$. For $z = 1$ he asks Alice and Bob for their choices while for $z = 2$ he measures Alice's and Bob's labs jointly in an entangled PBR measurement. The only events unavailable after the protocol are the 'free' choices a, b of Alice and Bob.

We next derive a contradiction of this scenario with the Local Friendliness assumptions. Regarding Absoluteness of Observed Events, the only two events unavailable after the protocol are Alice's and Bob's free choices. All other measurement outcomes and choice settings are available after the protocol. As a result, we only need the assumption that the freely chosen settings a, b of Alice and Bob, even if unavailable after the protocol, are absolute and single-valued, alongside the outcomes and choices available after the protocol.

Assumption 9 (Absoluteness of Observed Events (AOE))**.** An observed event is an absolute single event, not relative to anything or anyone.

Assumption 10 (Local Agency)**.** The only relevant events correlated with an intervention are in its future light cone.

A violation of Local Friendliness is then obtained as follows. If $a = 0, b = 0$ and $x = 1, y = 1$, then the choices of Alice and Bob are revealed by Charlie and Debbie[4] and the probabilities of Wigner for $z = 2$ are those of performing the PBR measurement on the state $|0\rangle \otimes |0\rangle$, finding that[5]

$$\wp(a{=}0, b{=}0, w{=}\xi_1 | x{=}1, y{=}1, z{=}2) = 0. \tag{5.4}$$

If $x = 1, y = 2$ and $a = 0, d = +$, then Charlie reveals Alice's choice by measuring the qubit T_A she received from Alice in the computational basis, which turns out to be $a = 0$, while Bob's choice is not revealed by Debbie who measured T_B in the X-basis obtaining $d = +$. Therefore, if $x = 1, y = 2, d = +$ and $a = 0$, then the probabilities of Wigner for $z = 2$ are those of performing the PBR measurement on $|0\rangle \otimes |+\rangle$, finding that

$$\wp(a{=}0, w{=}\xi_2 | x{=}1, y{=}2, d{=}+, z{=}2) = 0. \tag{5.5}$$

Similarly, if $b = 0$ and $x = 2, y = 1, z = 2, c = +$, the probabilities of Wigner are those of performing the PBR measurement on $|+\rangle \otimes |0\rangle$, finding that

$$\wp(b{=}0, w{=}\xi_3 | x{=}1, y{=}2, c{=}+, z{=}2) = 0. \tag{5.6}$$

Finally, if $x = 2, y = 2, z = 2, c = +, d = +$, then the probabilities of Wigner are those of performing the PBR measurement on the $|+\rangle \otimes |+\rangle$ state, finding

$$\wp(w{=}\xi_4 | x{=}2, y{=}2, c{=}+, d{=}+, z{=}2) = 0. \tag{5.7}$$

Next, we will use the Local Friendliness assumptions. In this case, AOE and Local Agency imply that $p(a, b, w|x, y, z) = p(a, b, w|z)$, i.e. the joint correlations of a, b, w, which always exist by AOE, should be independent of x and y. Moreover, they also imply that $p(a, b, w, c|x, y, z) = p(a, b, w, c|x, z)$ and $p(a, b, w, d|x, y, z) = p(a, b, w, d|y, z)$.

Applying AOE and Local Agency to Eq. (5.4), we obtain that

$$p(a{=}0, b{=}0, w{=}\xi_1 | x{=}2, y{=}2, z{=}2) = 0, \tag{5.8}$$

[4] Here we use a weak 'tracking assumption', see Ref. [27].

[5] In this section, we denote empirical probabilities by $\wp$, to distinguish them from probabilities p which are required to exist by AOE.

and, using $p(c{=}+, d{=}+|x{=}2, y{=}2, z{=}2) > 0$, that

$$p(a{=}0, b{=}0, w{=}\xi_1|x{=}2, y{=}2, c{=}+, d{=}+, z{=}2) = 0. \tag{5.9}$$

Similarly, applying AOE and Local Agency to Eq. (5.5), we obtain that

$$p(a{=}0, w{=}\xi_2|x{=}2, y{=}2, d{=}+, z{=}2) = 0, \tag{5.10}$$

and, using $p(c{=}+, d{=}+|x{=}2, y{=}2, z{=}2) > 0$, that

$$p(a{=}0, b{=}0, w{=}\xi_2|x{=}2, y{=}2, c{=}+, d{=}+, z{=}2) = 0. \tag{5.11}$$

The same reasoning for Eq. (5.6) leads to

$$p(a{=}0, b{=}0, w{=}\xi_3|x{=}2, y{=}2, c{=}+, d{=}+, z{=}2) = 0. \tag{5.12}$$

Summarizing, we have the following four conditions (from Eqs. (5.7), (5.9), (5.11) and (5.12)) on $p(a, b, w|E)$ with $E = (x{=}2, y{=}2, c{=}+, d{=}+, z{=}2)$:

$$\begin{aligned} p(a = 0, b = 0, w = \xi_1|E) = 0, \\ p(a = 0, b = 0, w = \xi_2|E) = 0, \\ p(a = 0, b = 0, w = \xi_3|E) = 0, \\ p(a = 0, b = 0, w = \xi_4|E) = 0. \end{aligned} \tag{5.13}$$

We thus find that

$$p(a{=}0, b{=}0, c{=}+, d{=}+|x{=}2, y{=}2, z{=}2) = 0. \tag{5.14}$$

However, this cannot be true, as by Local Agency we have that $p(a, b, c, d|x, y, z) = p(a, b, c, d|x, y)$ and

$$\wp(a{=}0, b{=}0, c{=}+, d{=}+|x{=}2, y{=}2, z{=}1) > 0. \tag{5.15}$$

We thus obtain a contradiction, finding a violation of the conjunction of Local Agency and the assumption that the events of Alice's and Bob's free choices, being the only unavailable values at the end of the protocol, are absolute and single-valued.

5.3 Generalizing to arbitrary choice amplitudes

In the protocol outlined above, we have assumed that Alice chooses $a = 0, 1$ with equal probabilities, or, more correctly, with equal amplitudes $1/\sqrt{2}$, leading to the state in Eq. (5.3) after her encoding of her choice in T_A, and similarly for Bob. But what if Alice's choice probabilities (and amplitudes) were different from $1/2$ (and $1/\sqrt{2}$)? In general, she could choose $a = 0, 1$ with arbitrary probabilities and amplitudes, leading to replacing Eq. (5.3) by

$$|\psi\rangle_{T_A L_A} = \alpha\,|00\rangle_{T_A L_A} + \beta\,|11\rangle_{T_A L_A}\,. \tag{5.16}$$

Then, to obtain a contradiction, we proceed as follows. We let Alice and Bob make their choices several times, while Charlie and Debbie perform state tomography on Alice's and Bob's labs when they have made their choices. This allows Charlie and Debbie to know α and β. Then, we run the protocol, and there are two possibilities.

The first possibility is that for $x = 1$, Charlie still decodes Alice's choice by measuring T_A in the computational basis, whereas for $x = 2$ he now measures T_A in a basis that contains the state

$$|\Gamma\rangle = \gamma\,|0\rangle + \delta\,|1\rangle\,, \tag{5.17}$$

with

$$\gamma^*\alpha = \delta^*\beta. \tag{5.18}$$

In this way, when Charlie obtains $c = \Gamma$, the post-selected state for L_A is $|+\rangle$, and similarly for Debbie choosing y, leading again to the relevant states $|00\rangle\,, |+\rangle\,|0\rangle\,, |0\rangle\,|+\rangle\,, |+\rangle\,|+\rangle$ of the PBR argument. Wigner acts as before, and a contradiction is obtained in the same way as in Section 5.2.

The second possibility is that Charlie and Debbie still measure in the Z- or X-basis, but that Wigner now performs a PBR measurement that discriminates between $|0\rangle$ and $\alpha\,|0\rangle + \beta\,|1\rangle$. For this, we actually need several additional agents to Alice and Bob and Charlie and Debbie. Namely, we need n agents making their choice and encoding their choice to send to some faraway observers who measure in the Z- or X-basis, with n determined by the general joint PBR measurement as in the original PBR argument [38].

5.4 On the modeling of Alice's and Bob's choices

In the above argument, regarding state tomography by Charlie and Debbie, we have assumed that each time that Alice and Bob make a choice which they encode in the computational bases of the qubits, the same coefficients α and β appear in Eq. (5.16). The motivation for this is that, assuming we initialized the state of Alice's lab each time in exactly the same state, and take the same exterior conditions, the unitary evolution of the lab is the same and always leads to Eq. (5.16). One could argue that there is some tension with the intuitive concept of a free choice in this case, indeed. Perhaps this suggests that one aspect of the free choice, the choice of the frequencies, only truly appears in open systems. Now, if we fed in different initial conditions for Alice's lab, a slightly different environment, then we would need Charlie and Debbie to additionally keep track of this environment control in order to obtain the coefficients α and β that depend on this environment. Then, in the actual protocol, when they choose their basis $|\Gamma\rangle, |\neg\Gamma\rangle$ they would need to additionally control on the environment states in order for the protocol to succeed.

5.5 Use of correlations

Note that our scenario can also be seen as Wigner swapping the entanglement between Alice and Charlie and Bob and Debbie to entanglement between Charlie and Debbie. The correlations used violate Bell nonlocality, as we show now.

The level of the contradiction in the Wigner's friend PBR argument is at the possibilistic level, so to check whether nonlocality fuels the contradiction, we need to check for nonlocality at the possibilistic level. We can see the obtained measurement outcomes in the Wigner's friend PBR argument as a tripartite Bell scenario, involving Charlie, Debbie and Wigner, choosing settings x, y, z and obtaining outcomes c, d, w. The initial state is given as in Eq. (5.2). Recall that Charlie chooses $x = 1, 2$ to measure T_A in the Z- or X-basis, and similarly for Debbie measuring T_B. Wigner makes a choice $z = 1, 2$. For $z = 2$ he performs the joint PBR measurement on $L_A \otimes L_B$ with outcome $w \in \{\xi_i\}_{i=1,\dots,4}$. For $z = 1$ he reveals Alice's and Bob's choices, meaning he measures $L_A \otimes L_B$ in the computational basis. Upon Wigner asking Alice and Bob for their outcome, let us label the corresponding outcome for $|ij\rangle_{L_A L_B}$ by $w = (i, j)|z = 1$ for $i, j \in \{0, 1\}$.

The empirical probabilities that are used in the scenario are

$$\begin{aligned}
&\wp(c{=}0, d{=}0, w{=}\xi_1 | x{=}1, y{=}1, z{=}2) = 0, \\
&\wp(c{=}0, d{=}+, w{=}\xi_2 | x{=}1, y{=}2, z{=}2) = 0, \\
&\wp(c{=}+, d{=}0, w{=}\xi_3 | x{=}2, y{=}1, z{=}2) = 0, \\
&\wp(c{=}+, d{=}+, w{=}\xi_4 | x{=}2, y{=}2, z{=}2) = 0,
\end{aligned} \tag{5.19}$$

and

$$\wp(c{=}+, d{=}+, w{=}(0,0) | x{=}2, y{=}2, z{=}1) > 0. \tag{5.20}$$

Moreover, to argue that $c|x = 1$ is actually Alice's choice a as revealed by Wigner when $z = 1$, and similarly for Debbie and Bob, we have that

$$\wp(c, d, w | x{=}1, y{=}1, z{=}1) = \delta_{w,(c,d)} \wp(w | z{=}1). \tag{5.21}$$

Now it is easy to see that these correlations are logically contextual (or, equivalently in this case, logically nonlocal, cf. Section 2.2.1). To simplify stating the argument, we label the outcome c when performing measurement $x = 1, 2$ (where for $x = 1$ Charlie measures Alice's qubit in the computational basis, decoding her choice a) by c_1, c_2, i.e., $c_x = (c|x)$ and similarly $d_y = (d|y)$ and $w_z = (w|z)$. Then, we find from Eq. (5.20) that the assignment $(c_2{=}+, d_2{=}+, w_1{=}(0,0))$ is possible in the context $C = (x{=}2, y{=}2, z{=}1)$. We now prove that there exists no global assignment extending this possible assignment $(c_2{=}+, d_2{=}+, w_1{=}(0,0))$, meaning the empirical correlations are logically contextual. By Eq. (5.21), $w_1 = (0,0)$ implies that $c_1 = 0 = d_1$. However, this assignment cannot be extended to a global assignment; one cannot assign a value to w_2 consistent with Eq. (5.19). Namely, as $c_1 = 0 = d_1$, by the first equation in Eq. (5.19), we cannot have $w_2 = \xi_1$. As $c_1 = 0, d_2 = +$, we cannot have $w_2 = \xi_2$ by the second equation in Eq. (5.19). Finally, as $c_2 = +, d_1 = 0$, by the third equation in Eq. (5.19) we cannot have $w_2 = \xi_3$, and the fourth equation and $c_2 = + = d_2$ imply that $w_2 \neq \xi_4$.

5.6 Discussion

We have provided a scenario that violates the Local Friendliness assumptions based on a scenario combining Wigner's friend with the PBR theorem. In fact, the only relevant variables that are unavailable after the protocol, are Alice's and Bob's free choices, and thus, under the same locality assumption of the Local Friendliness no-go theorem, this

provides an argument against the absoluteness of free choices. In order to do so, we had to model the free choices being made by Alice and Bob in a particular way. We have tried not to subscribe to any interpretation of free choices a priori, i.e. whether free choice is an emergent concept or not, and our argument is not an argument against the notion of free choice, only against the absoluteness of it, modulo quantum universality and a locality assumption. It might feel unnatural to model free choices in the way we did, and perhaps, in any way other than from the perspective of the entity who made that choice. In case one objects to such modeling, at least, our argument invites a careful operational rethinking of free choices and their role in physics.

— 6 —

Intermezzo: the Sleeping Beauty problem

In Chapter 3 we briefly encountered the Sleeping Beauty problem. In the Sleeping Beauty problem, Sleeping Beauty is put asleep, and a coin is tossed. For Tails, Sleeping Beauty is woken up on Monday, put asleep again and woken up again on Tuesday (with no recollection of her awakening on Monday). For Heads, Sleeping Beauty is woken up on Monday only. Each time Sleeping Beauty is woken up, she is asked what her belief is that the outcome was Heads. What should Sleeping Beauty reply? In literature arguments have been made for both $1/3$ and $1/2$ as the correct answer. In this chapter, I argue that $1/3$ is the correct answer, and not $1/2$, as Beauty being awake is nontrivial information that updates her beliefs about Heads to $1/3$. This claim is strengthened by considering an additional observer, Prince Probability, who may or may not meet Sleeping Beauty. If he meets an awake Sleeping Beauty, he lowers his credence in Heads to $1/3$.

Moreover, as argued in Chapter 3, the Sleeping Beauty problem serves as a classical example that violates FR's consistency condition, and thereby raises the question of what is genuinely *quantum* or *nonclassical* about the FR paradox [292]. Indeed, as we will see, Beauty and the Prince have different (probability) assignments. However, our refined version of the FR paradox, presented in Chapter 3, requires a weaker consistency condition, namely that observers need only agree when they meet and communicate about their thoughts. This condition is not violated in the Sleeping Beauty problem: when Beauty and the Prince meet, they will agree on their assignments.

In the remainder of this chapter, we discuss an operational argument that explains how the probabilities $1/2$ and $1/3$ arise, identify when each applies, and briefly relate to literature. This chapter is based on Ref. [120].

6.1 Introduction

The Sleeping Beauty problem I follow Ref. [292, Appendix C] for the description of the Sleeping Beauty[1] problem[2].

> *The Sleeping Beauty problem:* Sleeping Beauty participates in an experiment lasting 48 hours in which she is put to sleep on Sunday, and a fair coin is tossed. If the outcome is Heads, she will be woken up on Monday only and put back to sleep. If the outcome is Tails, she will be woken up on Monday and Tuesday. More specifically, for outcome Tails Sleeping Beauty is woken up on Monday and put back to sleep with her memory of her awakening erased, to be woken up on Tuesday again. Sleeping Beauty cannot distinguish between any of these awakenings. When woken up, Sleeping Beauty is asked for her credence that the outcome of the coin toss was Heads.

As a proposal to make the question somewhat clearer, we may ask Sleeping Beauty, when she is woken up, the following question: 'What do you now think the outcome was, Heads or Tails? And what is your credence now for Heads?'

Two possibilities for Sleeping Beauty to answer have been debated heavily in literature. The Halfers argue that $1/2$ is the correct answer, the main argument being the absence of any new information for Beauty when she wakes up so that she should stick to her credence $1/2$ she had on Sunday, whereas the Thirders argue that $1/3$ is the correct answer. Many works in favour of Elga's argument for $1/3$ [360], and against Lewis's $1/2$ arguments [363] have appeared [360, 364–379]. The arguments for $1/2$ mostly claim that no extra information is present for Beauty when she is woken up.[3] Here we argue that this claim is incorrect, in Section 6.2, using simple Bayesian probability theory and a slight extension of the protocol, with an outsider Prince Probability who may or may not meet Sleeping Beauty. We also consider the credence of a Sleeping Beauty who knows that she is dreaming. We believe that the strength of our argument is its simplicity. Of particular relevance is Rosenthal's argument [379], which gives an elementary conditional-probability argument for $1/3$ by introducing an auxiliary dime

[1] We will sometimes abbreviate Sleeping Beauty as SB or Beauty.

[2] The SB problem was first written down by Elga in [360], who attributed it to Robert Stalnaker and Arnold Zuboff. Similar problems were posed in [361, 362].

[3] The fact that this is false was already noted (among others) in [365, 370, 371]. In this chapter we give further arguments for this claim, by considering a concrete operationalization of the problem and considering the viewpoint of an outsider who may or may not meet Sleeping Beauty.

in the Sleeping Beauty experiment. Our approach is related, but operationalizes the relevant evidence as Beauty's being awake now and makes the corresponding update explicit from the perspective of an external observer. We consider variations of the SB problem in Section 6.3, and relate to Elga's argument and literature in Section 6.4.

6.2 A Thirder's Tale of Beauty and the Prince

In this section we will argue how Sleeping Beauty knowing that she is awake is useful, nontrivial information for Sleeping Beauty, leading to the Thirder position. First, we consider Beauty's point of view, but perhaps most convincing is the story including an outsider Prince Probability, who may or may not meet Beauty and updates his credence in Heads accordingly, which we discuss subsequently. Finally, we argue what credence Beauty would assign if she knew she were dreaming (and thus asleep), the same credence that Prince would assign upon encountering Sleeping Beauty whilst she is asleep.

Beauty knowing that she is awake is nontrivial information. In the Sleeping Beauty protocol, Beauty is asked to give her credence for Heads when being woken up. Importantly, here the fact that Beauty is awake is nontrivial: she could have been in a state of sleep too. Suppose for example that each time Beauty is woken up she is awake for z hours (with $z < 24$). The event of her being awake under Heads then has probability

$$p(\text{SB awake}|H) = z/(2 \cdot 24) = \frac{1}{2}x, \tag{6.1}$$

with $x = z/24$, as she is awake for only z hours in the two days (48 hours) of the experiment.[4] On the other hand, under Tails we have

$$p(\text{SB awake}|T) = \frac{2z}{48} = x. \tag{6.2}$$

The probability of Beauty being awake is thus

$$p(\text{SB awake}) = p(H)p(\text{SB awake}|H) + p(T)p(\text{SB awake}|T) = \frac{1}{2} \cdot \frac{1}{2}x + \frac{1}{2} \cdot x = \frac{3}{4}x. \tag{6.3}$$

[4]Here and further on, the events H and T stand for the outcome of the coin toss being Heads and Tails, respectively.

Therefore, the credence Beauty gives to Heads while being awake is

$$p(H|\text{SB awake}) = \frac{p(H)p(\text{SB awake}|H)}{p(\text{SB awake})} = \frac{1/2 \cdot 1/2x}{3/4x} = 1/3. \tag{6.4}$$

Analogously, we find that $p(T|\text{SB awake}) = 2/3$. Thus indeed, while being awake Sleeping Beauty gives credence $1/3$ to Heads, using the fact that she is awake, because she is more likely to be awake when the outcome is Tails.

This argument can be easily extended to an experiment lasting N days, where Beauty is woken up on N consecutive days for Tails, whereas only once for Heads. In that case, when woken up, Beauty's belief in Heads is $1/(N+1)$. Furthermore, if the coin toss is also unfair, with $P(H) = h$, it is easy to verify that our method that led to Eq. (6.4) gives that $P(H|\text{SB awake}) = \frac{1}{N \cdot \frac{1-h}{h} + 1}$.

Perspective of outsider Prince Probability. Consider an outsider, Prince Probability[5], and for concreteness let Sleeping Beauty always reside in a sealed lab. When not partaking in the protocol, Prince Probability assigns credence $1/2$ to Tails. Prince Probability has no idea of which day it is, and at any instance Prince Probability can choose to visit Beauty's lab. If he does so, he sees Beauty either asleep or awake. If Prince Probability sees Beauty being awake, he will lower his credence to $1/3$ for Heads because under Tails she is awake for more time. The calculation is analogous to the one right above in Eq. (6.4). Upon meeting the Prince, Beauty does not seem to get any relevant information for Heads versus Tails, suggesting that she should indeed always assign credence $1/3$ to Heads when she is awake. We have sketched this protocol in Fig. 6.1.

This is in line with arguments in literature by Neal, Stalnaker and Horgan [375, 376, 380]. They considered an additional agent also part of the experiment (called Prince, Sleeping Ugly and Sleeping Betty, respectively), who was woken up both on Monday and Tuesday. These arguments are similar to ours, with the difference that our Prince Probability is truly an outsider to the experiment, not put to sleep.

What if Sleeping Beauty knew she was dreaming? Consider the Sleeping Beauty problem, but now Beauty knows she is dreaming (and thus asleep, assuming she always dreams while asleep). Using similar Bayesian arguments as above, she will now increase her credence in Heads. Suppose Sleeping Beauty is awake for x hours each time she

[5]We will also refer to Prince Probability simply as Prince or PP.

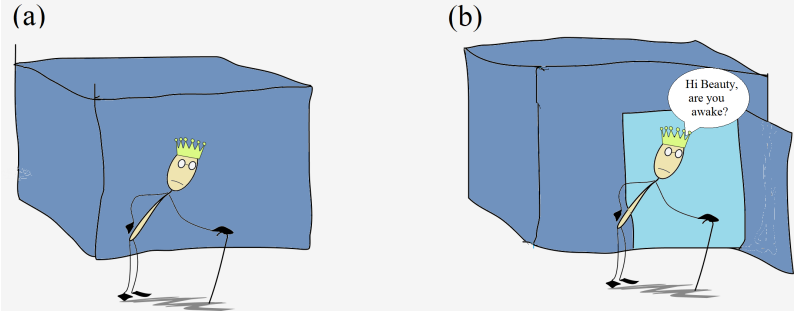

Figure 6.1: A schematic overview of the extension of the Sleeping Beauty problem that includes an outsider's perspective (Prince Probability's) as described in Section 6.2. (a): Sleeping Beauty resides in a sealed lab and an outsider Prince Probability assigns a probability $1/2$ to Heads. (b): Prince Probability opens Sleeping Beauty's lab and can now see whether she is awake or asleep. If she is awake, Prince Probability lowers his credence for Heads to $1/3$. If she is asleep, he increases his credence for Heads.

is woken up (so $x < 24$ and we assume x to be known to her). We have that $P(\text{SB asleep}|H) = (48-x)/48 = 1-y$ and $P(\text{SB asleep}|T) = (48-2x)/48 = 1-2y$ where we have defined $y = x/48$ (with now $y < 1/2$). Therefore,

$$\begin{aligned} P(\text{SB asleep}) &= P(H)P(\text{SB asleep}|H) + P(T)P(\text{SB asleep}|T) \\ &= 1/2 \cdot (1-y) + 1/2 \cdot (1-2y) = 1/2 \cdot (2-3y), \end{aligned} \tag{6.5}$$

and thus

$$P(H|\text{SB asleep}) = \frac{P(H)P(\text{SB asleep}|H)}{P(\text{SB asleep})} = \frac{1/2 \cdot (1-y)}{1/2 \cdot (2-3y)} = \frac{1-y}{2-3y}. \tag{6.6}$$

Thus, knowing that she is currently dreaming, Sleeping Beauty assigns credence $(1-y)/(2-3y)$ to Heads in this scenario. This is also the credence an outsider Prince Probability would give to Heads when he sees that SB is asleep. Importantly, through y this result does depend on the time span Sleeping Beauty is awake for each time when woken up. For example, if each time she wakes up, SB is awake for 23 (1) hours, she gives credence 0.926... (0.505...) to Heads while knowing that she is dreaming. This difference and dependence on y make sense intuitively: for $x = 23$ Beauty is asleep for 25 hours if Heads and only 2 hours if Tails, whereas for $x = 1$ Beauty is asleep for 47 hours if Heads and for 46 hours if Tails, a much smaller difference.

6.3 A variation of the Sleeping Beauty problem

We now discuss a slightly altered version of the copy-SB version of Ref. [292], similar to the Sailor's Child problem [375] and Rosenthal's sleeping-twins variant [379], which has a resolution similar to the one we proposed above in Section 6.2 for the original SB problem.

The copy-SB protocol There are two identical sealed labs, indistinguishable from the outside and inside. Sleeping Beauty resides in one of those labs. A fair coin is tossed, with the outcome not revealed to Sleeping Beauty. If the outcome is Tails, a copy of Sleeping Beauty is made and put in the second lab, so that in that case both labs contain a Sleeping Beauty. Again Sleeping Beauty knows the protocol. Each version of Sleeping Beauty is asked whether she thinks the outcome of the coin toss was Heads or Tails, and what her credence for Heads is. We will consider again the perspective of an outsider Prince Probability. All Prince Probability knows is the protocol: he knows that there are two labs, but he does not know in which lab the original Sleeping Beauty resided. He knows there was a coin toss that may have led to the copying of Sleeping Beauty so that both labs would contain a Sleeping Beauty, but he does not know the outcome of the coin toss, i.e. he does not know whether both labs contain a Sleeping Beauty or only one and which lab contains a Sleeping Beauty.

Resolution of the copy-SB problem We will argue that if SB is woken up and is asked which outcome the coin toss had, she would assign $1/3$ probability for Heads. An outsider, Prince Probability, who does not know the outcome of the coin toss, gives credence $1/2$ to Heads prior to any action he undertakes. But when he meets a SB, he changes his credence to $1/3$ for Heads. Therefore, when Prince Probability and Beauty meet, they agree on their probability assignments for Heads, namely $1/3$. Before they have met, however, Prince Probability and Beauty can each correctly assign different credences to Heads, namely $1/2$ by Prince Probability and $1/3$ by Beauty.

Prince Probability's point of view The probability that opening a lab Prince Probability finds Beauty is

$$\begin{aligned} p(\text{PP finds SB}) &= p(T)p(\text{PP finds SB}|T) + p(H)p(\text{PP finds SB}|H) \\ &= \frac{1}{2} \cdot 1 + \frac{1}{2}\frac{1}{2} = \frac{3}{4} \end{aligned} \tag{6.7}$$

and similarly, the probability of Prince Probability finding a Beauty upon opening the door of a lab under outcome Heads (which we refer to as finding a Heads SB) is $p(\text{PP finds SB} \wedge H) = \frac{1}{4}$. Therefore, if Prince Probability finds a Sleeping Beauty, the probability that the outcome of the coin toss was Heads is equal to $p(\text{PP finds SB} \wedge H|\text{PP finds SB}) = \frac{1}{3}$.

Thus, before Prince Probability enters a lab, he argues that the probability for Heads is $1/2$. But, when he enters a lab, he changes his credence in Heads, depending on whether or not there is a Beauty. Namely, if he finds a Beauty, his credence in Heads becomes $1/3$. This is not surprising: Prince Probability has gotten new information, namely that he sees a Beauty. For outcome Tails he is more likely to find a Beauty. Thus, conditioned on Prince Probability finding a Beauty, he gives credence $1/3$ to Heads. Still, Prince Probability assigns probability $1/2$ to Heads before he undertakes any action.

What if Prince Probability finds no Sleeping Beauty? If there are only two labs and if Prince Probability opens a lab and finds no Beauty, he will know the outcome was Heads for sure. If there were additional empty labs (so under Tails only still two labs contain a Beauty), then Prince Probability will have credence higher than $1/2$ but less than 1 that the outcome was Heads, as he is more likely to find no Beauty if the outcome were Heads. His credence here thus depends on the number of labs there are, similar to the discussion of Prince Probability finding Sleeping Beauty asleep in Section 6.2.

Sleeping Beauty's point of view From the above discussion, we found that when Prince Probability and a Beauty meet, Prince Probability assigns probability $1/3$ to Heads. But from Beauty's point of view, Prince Probability visiting her does not give her useful information: Prince Probability simply chooses to enter her lab or the other one, with equal probability. This seems to suggest that Beauty indeed always assigns probability $1/3$. Namely, Beauty names the two labs Lab 1 and Lab 2, and without loss of generality assumes she is in Lab 2. So the fact that she exists is the event 'There is a SB in Lab 2' or shortly 'SB in Lab 2'. However, while outside the labs, Prince Probability can name the two labs 1 and 2 to his liking, but he cannot make a statement such as 'There is a

SB in Lab 2', he can only say 'Lab 1 or Lab 2 contain a SB'. Thus, Sleeping Beauty's credence (of $1/3$) in Heads is conditional on the event 'SB in Lab 2', a statement that Prince Probability having credence $1/2$ before entering the labs cannot make. This is an almost trivial observation for Sleeping Beauty, but one that matters. Namely, if the outcome of the coin toss is Tails, the probability of 'SB in Lab 2' is given by

$$p(\text{SB in Lab 2}|T) = 1, \tag{6.8}$$

whereas for Heads there is a Beauty either in Lab 1 or 2, but not both,

$$p(\text{SB in Lab 2}|H) = \frac{1}{2}. \tag{6.9}$$

We also have that

$$p(\text{SB in Lab 2}) = p(T)p(\text{SB in Lab 2}|T) + p(H)p(\text{SB in Lab 2}|H) = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot\frac{1}{2} = \frac{3}{4}. \tag{6.10}$$

Therefore, the credence for Beauty for Tails is

$$p(T|\text{SB in Lab 2}) = \frac{p(T)p(\text{SB in Lab 2})}{p(\text{SB in Lab 2})} = \frac{1/2\cdot 1}{3/4} = 2/3. \tag{6.11}$$

The credence Sleeping Beauty gives to Heads is thus $1/3$, where she uses the fact that she exists (i.e. Lab 2 contains her).

6.4 Relating to literature on Sleeping Beauty

Elga's argument using the Principle of Indifference for self-locating belief leads to the same conclusion as ours in Section 6.2, namely that Sleeping Beauty should assign credence $1/3$ to Heads when woken up [360]. Furthermore, also in variations of the protocol, such as an unfair coin toss and many awakenings, Elga's resolution method gives the same answer as ours, and accords with the updating of the outsider Prince Probability's point of view when meeting a Sleeping Beauty who is awake. We can see this fact as an argument for the correctness of Elga's reasoning.

After repeating Elga's argument and showing how it agrees with our arguments, we briefly discuss what arguments on the Sleeping Beauty problem have been given in literature, and counter some of the prominent Halfer arguments.

6.4.1 Elga's argument for 1/3 and the Principle of Indifference

Elga [360] uses the Principle of Indifference for self-locating belief (see also Ref. [364]) to argue that the correct answer for Sleeping Beauty to give when woken up is $1/3$. More precisely, in Ref. [364] Elga states the Principle of Indifference for self-locating belief as follows (quoting from Ref. [364]):

> Principle of Indifference: Similar centered worlds deserve equal credence.
>
> If Sleeping Beauty is woken up, she knows she is in one of three events:
>
> H_1: Heads and Monday,
>
> T_1: Tails and Monday,
>
> T_2: Tails and Tuesday.

Elga's resolution then proceeds by arguing that $P(T_1) = P(T_2)$ and $P(T_1) = P(H_1)$. Namely, given Tails, T_1 and T_2 take place in similar centered worlds, so by the Principle of Indifference for self-locating belief $P(T_1|T_1 \text{ or } T_2) = P(T_2|T_1 \text{ or } T_2)$, so that $P(T_1) = P(T_2)$. Alternatively to the Sleeping Beauty problem where the coin is flipped *first* on Sunday, we could also imagine the protocol where Sleeping Beauty is woken up on Monday and *then* a coin is flipped determining whether Sleeping Beauty is also woken up on Tuesday (Tails) or not (Heads). Elga argues that in both cases, if Sleeping Beauty were told that it is Monday, she must assign the same credence to Heads, namely $1/2$, i.e. $P(H_1|H_1 \text{ or } T_1) = 1/2$, leading to $P(H_1) = P(T_1)$. Using $P(T_1) = P(T_2)$ and $P(H_1) = P(T_1)$ together with the normalisation $P(T_1) + P(T_2) + P(H_1) = 1$, we obtain that indeed $P(H) = P(H_1) = 1/3$. Elga thus argues that when woken up Sleeping Beauty should give credence $1/3$ to Heads, in agreement with our resolution.

Importantly, the Principle of Indifference does *not* allow you to argue that, as T_1, T_2, H_1 are all subjectively indistinguishable, they should all get equal credence $1/3$. For example, in the case of an unfair coin toss, Elga's argument will correctly divert from credence $1/3$ for Heads. For example, if the unfair coin toss gives Heads with probability $h = 1/4$, Elga's argument goes as follows. Still, T_1 and T_2 take place in similar centered worlds, so $P(T_1) = P(T_2)$. Suppose it is Monday, so we condition on the event H_1 or T_1 (SB waking up on Monday). Then, by the same argument as above, we could also take the coin to be tossed after Sleeping Beauty waking up on Monday and being put back to sleep, for which then Heads has probability $h = 1/4$. So $P(H_1|H_1 \text{ or } T_1) = h$, giving $P(H_1) = hP(H_1) + hP(T_1)$, and thus $P(H_1) = h/(1-h)P(T_1)$. Combining this with

$P(T_1) = P(T_2)$ and the normalisation $P(H_1) + P(T_1) + P(T_2) = 1$, we obtain, for example for $h = 1/4$, that $P(H_1) = 1/7$. It is easy to verify that also in the general case of N awakenings and an unfair coin toss with $P(H) = h$, Elga's resolution accords with ours and Rosenthal's [379], namely $P(H|\text{SB awake}) = \frac{1}{N \cdot \frac{1-h}{h} + 1}$.

As noted by Elga himself using an entertaining story involving Dr Evil [364], a consequence of the Principle of Indifference that seems somewhat unsettling is the fact that it favors situations where there are many duplicates of ourselves, see also Ref. [381]. Similar difficulties arise when applying the Principle of Indifference to the many-worlds interpretation of quantum mechanics [381] and Boltzmann Brain universes[6] in cosmology [292, 381]. We will not explore these interesting questions further here.

6.4.2 The Thirders vs. the Halfers

Lewis argues in [363] against Elga's resolution of $1/3$, and favors $1/2$, based on the apparent (but false) idea that no new relevant information is gained by Sleeping Beauty waking up. In Section 6.2 we have countered this thought by operationalising the SB scenario somewhat more. In that case it is clear that by knowing that she is awake, Beauty does obtain relevant new information, i.e. Beauty being woken up is a nontrivial fact during the two days of the experiment. The fact that this piece of information is nontrivial, namely 'I (SB) am awake *now*', has also been noted in (among others) Refs. [365, 370, 371], as an argument against $1/2$. Monton argued for a change in information too, namely that when Beauty wakes up compared to before she was put to sleep on Sunday Beauty has a loss of her temporal location [372]. Moreover, using a variation of the protocol, Dorr [368] and Arntzenius [367] argued how indeed at Beauty's awakening there is nontrivial new information for her. Rosenthal [379] uses conditional probabilities while introducing an additional dime to argue for $1/3$; his arguments and results are in line with ours, and his derivation may be viewed as complementary to ours. Our more operational setting and introduction of the outsider's view Prince Probability who may or may not meet Beauty thus provide strong support for these arguments.

White provided a generalized Sleeping Beauty problem in Ref. [383] that, in his view,

[6] Another uneasiness about Boltzmann Brains is the following. Boltzmann Brains arise in our models of physics and cosmology, but Boltzmann Brains have untrustworthy memories, i.e. their memories of 'observed' data are not in line with the universe they live in [382]. Thus if we were Boltzmann Brains, we should not trust our memories of 'observed' data, undermining the use of our physical theories and thus the argument for the existence of Boltzmann Brains.

seems to argue for $1/2$. However, for his argument in favor of $1/2$ using this generalized Sleeping Beauty problem, White uses a probability conditioned on the event 'SB is being woken up at least once', which is different from the self-locating event where Sleeping Beauty notices that she has been woken up, also noted in Refs. [369, 384]. Namely, before Prince Probability enters any lab (and thus assigns credence $1/2$ to Heads), he obviously knows that Beauty is woken up at least once. It is only when he actually enters a lab and meets Beauty while she is awake that he updates his credence in Heads to $1/3$. The conditioned event $W =$ 'SB awake at least once' is different from 'SB awake *now*', corresponding to the statements 'SB in Lab 1 or 2' versus 'SB in Lab 2' of Section 6.3.

The Bayesian rule of updating (conditionalization) for self-locating beliefs has also been questioned in favor of $1/2$ answers for the SB problem in Ref. [385]. However, having given arguments in favor of $1/3$, especially our argument with the outsider Prince Probability who when meeting an awake SB changes his credence to $1/3$ for Heads, questions such proposed change for the Bayesian update rule for self-locating beliefs as in Ref. [385].

For more literature on the Sleeping Beauty problem, we refer the reader to Ref. [369, 381].

6.5 Conclusion

In this chapter we have given arguments employing simple Bayesian probability that suggest $1/3$ to be the correct answer for Sleeping Beauty to reply when she is woken up and asked for her credence in Heads. Sleeping Beauty does have extra useful information, which is why $1/2$ is not her correct answer. An outsider, Prince Probability, assigns credence $1/2$ to Heads, but when Prince Probability meets Sleeping Beauty being awake, he lowers his credence in Heads to $1/3$. Furthermore, if Prince Probability meets Sleeping Beauty while asleep, he instead increases his credence in Heads.

Part II

Universality and black holes

The central assumption of Part I was that quantum theory applies equally to microscopic laboratory systems, measurement devices and macroscopic observers. Under this hypothesis, we explored several foundational results that challenge our naive understanding of reality. We now turn to another question of universality. How can quantum theory, assuming its universal applicability extends to gravity, resolve the puzzles that black holes pose? While we will not fully answer this question, we extend these paradoxes further, point out subtleties concerning some standard assumptions, and present a calculation in which aspects of quantum theory are combined with general relativity in the form of a superposed black hole.

After briefly outlining black hole paradoxes and proposed resolutions in literature in Chapter 7, the remainder of this part is organized as follows. Bridging Part I and Part II of this thesis, we present several scenarios in Chapter 8 that merge extended Wigner's friend arguments with black hole physics. These scenarios yield paradoxes that strengthen the cloning and firewall paradoxes, and rely on standard assumptions of black hole puzzles. We highlight several important subtleties in these assumptions, inviting a more careful verification thereof towards resolving these puzzles. Among the methods that can be employed for this purpose are the use of quantum reference frames and operational investigations of a black hole and its radiation. In Chapter 9, we illustrate the combination of these approaches by analyzing the response of a particle detector near a position-superposed black hole using aspects of quantum reference frames. This analysis was motivated by the study of a detector's response near a mass-superposed black hole in literature [119]. Chapters 8 and 9 are based on Refs. [386] and [387], respectively.

We end with a short discussion and conclusion of both parts in Chapter 10.

— 7 —

A brief introduction to black hole puzzles

Arguably, few phenomena in the universe are as captivating as black holes, where two of the most foundational theories of modern science – general relativity and quantum theory – intersect in profound ways. Applying principles of quantum theory to black holes, Hawking proved in his seminal paper [45] that black holes radiate, and by energy conservation thereby evaporate over time. However, Hawking's calculations suggested this radiation to be approximately thermal, which seemed to indicate a loss of information: an initial pure state of matter collapsing to a black hole would evolve nonunitarily to a mixed state of radiation after evaporation [46]. This enigma is also referred to as the black hole information puzzle [49, 388].

Page gave a general argument of how a black hole system might evolve, based on the assumption that the black hole and radiation combined are always in a random pure state [62, 389]. The resulting curve for the (renormalised [49, 390]) entanglement entropy[1] is estimated to decrease somewhat past halfway of the evaporation [62, 63], where it diverges from the entanglement entropy per Hawking's calculation.

From Page's unitary entropy curve, late-time radiation should be highly entangled with the early radiation. Moreover, the equivalence principle and quantum field theory (QFT) in curved spacetime suggest high entanglement across the horizon [392]. Therefore, late-time radiation just outside the black hole should be highly entangled with both the early radiation and modes inside the black hole, apparently violating monogamy of entanglement, also known as the firewall paradox [50–52, 393].

Another black hole puzzle is a no-cloning argument analyzed by Susskind and Thor-

[1]We generally distinguish two types of entropy: the entanglement entropy, also referred to as the von Neumann or fine-grained entropy, describing the entanglement of one system with another, and the thermodynamic entropy, a coarse-graining thereof [391].

lacius [47] and later by Hayden and Preskill [48], suggesting a form of black hole complementarity [47, 72, 394]. The firewall argument is sometimes presented as an argument against complementarity [49], but additionally assumes the approximate validity of semiclassical gravity to argue for high entanglement across the horizon.

In the rest of this chapter, we briefly outline arguments in literature arguing for unitarity, and comment upon an aspect overlooked by Hawking when originally formulating his information loss paradox, a black hole's soft hair.

7.1 Literature on the black hole information puzzle

The resolution of the black hole information puzzle remains heavily debated and inconclusive, but much work has been done since, hinting that information is preserved. Expectations that black hole evaporation is a unitary process come from AdS/CFT [49, 74], and more recently from the proposed existence of 'entanglement islands' inside the black hole giving a Page curve for the radiation [68, 75–94].[2] Replica trick calculations of entanglement entropy in 1+1d Euclidean gravity [395] suggest that additional saddle points —– so-called replica wormholes [396–398]—appear when entanglement islands form. Such replica wormholes lack a clear interpretation, but have been proposed to relate to black hole hair [399, 400]. Unitarity has also been argued for from two-dimensional investigations [401–404].

Another proposal for information recovery consists of stable or long-lived remnants, objects that look small from the outside but store an enormous amount of information in their interior [405, 406]. An example of such proposal is black-to-white hole tunneling [407–414]. Arguments in favor of remnants argue that there is a notion of volume that is very large for old black holes [415, 416]. Remnants are, however, often argued to encounter problems such as infinite pair production [417–419][3]. A related idea consists of baby universes that form inside black holes [420, 421]. If a (Planck-scale) remnant held the information, there would be no need for the Hawking radiation to purify (until the remnant potentially decays and spits out the information it holds in low-energetic radiation), and thus also no Page curve for the (pre-remnant) Hawking radiation, no

[2] Concerns have been expressed that such Page curve from islands only arises in massive gravity [65, 68, 95, 96, 98, 99]; in turn, a reply can be found in Ref. [97], and a reply to this reply in Ref. [94]. See also footnote 15 in Chapter 8.

[3] For counterarguments see [405, 406, 410].

cloning and firewall paradoxes (and thus also no black hole Wigner's friend arguments of this work). We will not discuss this option further here.

Proposals that change the structure or existence of the horizon include, in addition to firewalls, fuzzballs [422], nonviolent nonlocality [423], gravastars as an alternative to black holes [424] and that evaporation may be halted earlier than expected leading to 'dark stars', which, however, by itself may not yield a full resolution to the information puzzle [425–430].

One recent idea regarding a physical mechanism for unitary black hole evaporation includes soft hair, which we discuss more in Section 7.2, low-energy gravitons dressing black hole spacetimes that could purify the Hawking radiation [101–106, 431]. Soft hair and soft theorems [432] are deeply tied to (gravitational) memory [433–437] and asymptotic symmetries [110, 111, 438–443] through the infrared triangle [102], and indeed argued to be of crucial importance for infrared problems in massless scattering [442–448]. Related proposals are quantum hair from higher-order curvature corrections [399, 431, 449, 450] and black-hole gravitational condensates [451], and a black hole wavefunction, which can be seen as a form of hair, that spreads or branches out [106, 399, 431, 452–455].

7.2 A black hole's soft hair

Near infinity, far away from the black hole, what a particle detector such as the Unruh–DeWitt one probes is the radiation density matrix $\rho_{w,w'}$ defined by [452, 455–457]

$$\rho_{w,w'} = \int_0^\infty \mathrm{d}\omega'' \beta_{ww''} \beta^*_{w'w''}, \tag{7.1}$$

with β the Bogoliubov coefficients relating past and future vacua and w shorthand for the frequency and angular labels (ω, l, m). This density matrix was suggested to be approximately thermal by Hawking [45, 456], being only non-zero for $\omega = \omega', l = l', m = m'$. Therefore, it cannot be pure, even if the black hole were formed from a pure state. As this calculation does not take into account any backreaction, this leaves a mechanism for early radiation to get entangled with the black hole's properties, and thereby also with the late radiation. When taking into account backreaction, conservation laws indeed suggest that the radiation's frequency and angular labels become entangled with the black hole mass and momentum, and thereby potentially influence later emissions. Yet, the no-hair theorems, or at least a naive reading thereof, tell us that the only relevant

properties of a black hole are its mass, position, angular and linear momentum, and electromagnetic charge. Therefore, the radiation's infinite set of angular momentum l-labels only has the 11 black hole no-hair degrees of freedom (in 4d) to correlate with, too few to purify its state. This (apparent) problem of nonunitary evolution is also referred to as the black hole information puzzle.

However, in the last 15 years it has become clear that there are potentially many more degrees of freedom of the black hole spacetime that the radiation can get entangled with, the black hole's soft hair [101, 103, 458]. In fact, already since the pioneering work of Bondi, van der Burg, Metzner and Sachs [438, 439], it has been known that the asymptotic symmetry group of asymptotically flat spacetimes (in 4d) is not just the Poincaré group of translations, rotations and boosts, but an infinite-dimensional group, enlarged with supertranslations, angle-dependent translations (and moreover potentially extended to encompass superrotations and superboosts [102, 110, 459, 460]). These additional symmetries endow the spacetime with many more nontrivial charges, and stem from diffeomorphisms that can locally be gauged away but act nontrivially at infinity. Such asymptotic symmetries are deeply tied to the memory effect [433, 434, 436, 461], the lasting displacement of the relative positions of detectors after the passing of a gravitational wave, and soft theorems [432] through the infrared triangle [102, 118, 462]. Related is a long-standing problem in massless quantum scattering, where the S-matrix was long known to suffer from infrared divergences [444–448]. Historically, this has been dealt with by calculating inclusive cross sections, focusing on observables agnostic to the infrared sector [432, 447, 463], but a complete, satisfactory treatment remains problematic. Incorporating memory effects appears crucial for developing such a framework [443]. The field of quantum reference frames [151, 329, 464–472] may have an enlightening role to play in this story [469, 470], with Goldstone modes labeling the degenerate vacua suggested to arise from the relation between internal and external reference frames [469]. Therefore, a black hole's soft hair was an aspect overlooked by Hawking when phrasing his argument for the information puzzle, and thus might play an important role in its resolution, or at least in a complete quantum treatment of a black hole's evolution.

7.3 What about paradoxes for infalling observers?

Despite these arguments for unitarity, a clear, robust physical understanding of black hole paradoxes for infalling observers such as the firewall and cloning paradoxes remains elusive. One proposal is black hole complementarity [47, 72, 73, 394, 473] where the inside of a black hole and its radiation outside are not independent, and can be motivated from nontrivial relations between infalling matter and outgoing modes by backreaction [73]. Complementarity is only a principle, for which the firewall paradox was originally presented as a challenge [52, 474], and can be formalized in different ways for an actual theory of the interior [49]. Modeling black hole evaporation through quantum circuits suggests the importance of computational complexity and quantum error correction [53, 475–479], where it has been argued that computational complexity protects the semiclassical causal structure of spacetime [477], i.e. computational complexity preventing someone from creating a 'firewall' of horizon excitations by manipulating the Hawking radiation.

However, even with complexity [49, 53, 475–477, 479, 480] and waiting time [48] arguments to avoid operational contradictions for the firewall and cloning paradoxes, quantum theory cannot give a single description of two quantum clones or a violation of monogamy of entanglement. Specific proposals for the interior include state-dependence [481–483], also argued for in a milder form in the paradigm of entanglement wedge reconstructions and islands [75, 484], final state projection [485, 486], strong complementarity [49] and proposals that modify aspects around the horizon, either violently such as firewalls, or (potentially) more subtly through nonviolent nonlocality [423, 487] and string-theoretic fuzzballs [422], see also Refs. [49, 429] for more discussion on other proposed compact quantum objects. Other proposals view the interior or spacetime as emergent concepts [451, 488], with Ref. [488] shown to exhibit a form of complementarity and state-dependence.

We will revisit these paradoxes for infalling observers in Chapter 8, more carefully analyzing their assumptions, and combining them with extended Wigner's friend paradoxes. As we will discuss in Section 8.4 of Chapter 8, important subtleties for paradoxes featuring infalling observers may be superposed quantum geometries and issues of locality and gauge-invariance.

— 8 —

Wigner's friend's black hole adventure

At the heart of both black hole physics and Wigner's friend scenarios lies the question of unitarity. In Wigner's friend setups, sealed-lab measurements are modeled unitarily, probing the measurement problem. In black hole physics, the unitarity problem concerns information preservation in evaporation. Inspired by recent work of Hausmann and Renner [107], who draw analogies between Wigner's friend and black hole paradoxes, we combine extensions of the Wigner's friend scenario with black hole physics to derive no-go theorems that strengthen the black hole no-cloning and firewall paradoxes. More specifically, the cloning paradox leaves room for a theory that could describe the entire black hole physics of the interior and exterior while satisfying the assumptions of these paradoxes but without operational contradictions concerning quantum theory's predictions. We close this loophole, showing that no such theory exists modulo the standard assumptions of the cloning paradox. Similarly, we lift the firewall argument to a paradox showing that no such theory exists when assuming persistent outcome records in the black hole interior. We also give an argument for why the firewall paradox assumptions give rise to a violation of no-signalling. We will employ standard assumptions of the cloning and firewall paradoxes for our protocols regarding unitarity and the decoding of Hawking radiation, but provide a more critical assessment in Section 8.4 by briefly outlining several important subtleties that question the validity of these assumptions.

The assumption that superobservers as in the Wigner's friend scenario exist can be argued to be not much more exotic than the capability of an observer to reconstruct the initial state of a complicated quantum system from its Hawking radiation. In our protocols and no-go theorems, we assume that the superobserver can perform arbitrary quantum operations on an observer and its environment, using the wording of Ref. [12]. This chapter focuses on 4-dimensional asymptotically flat Schwarzschild black holes, but

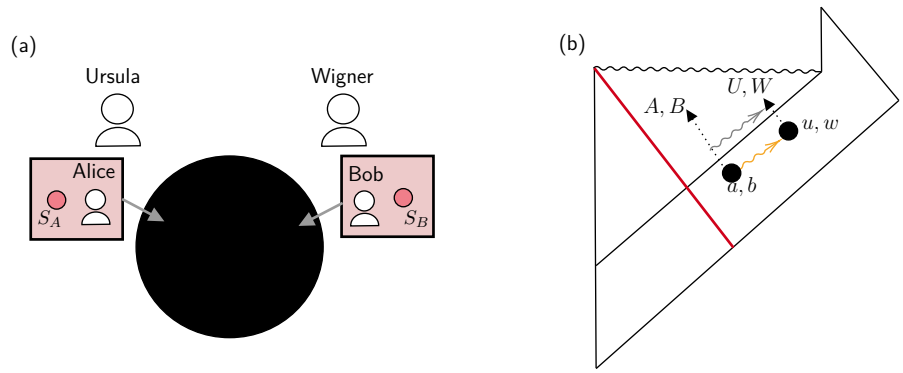

Figure 8.1: Initial set-up of our first protocol (a) and pictured in a Penrose diagram (b) for a black hole formed from collapse. Alice and Bob measure S_A, S_B and fall into the black hole. Ursula and Wigner catch their Hawking radiation (yellow wiggly arrow) upon which they perform an operation, after which they fall into the black hole to retrieve the outcomes of Alice and Bob (green wiggly arrow). As pointed out in Section 8.1.1 and discussed further in Section 8.4, one may question, among others, whether the location of the retrieval and decoding of Hawking radiation can be well into the bulk, whether Ursula and Wigner can still fall into the black hole after this decoding, and the correct causal structure. We also provide a theorem (Theorem 7) where Ursula and Wigner remain outside the black hole.

our arguments can be extended to other types of black holes.

The remainder of this chapter is organized as follows. We present our two main protocols in Sections 8.1 and 8.2, combining extended Wigner's friend paradoxes with the black hole cloning and firewall paradoxes. Next, we relate to the work of Hausmann and Renner [107] in Section 8.3, and conclude by highlighting some subtleties underlying black hole puzzles in Section 8.4.

8.1 Combining extended Wigner's friend paradoxes and black hole cloning

In this section, we construct a protocol that combines the Frauchiger–Renner and black hole cloning paradoxes and derive two no-go theorems. We briefly compare to the cloning paradox. Implications are discussed in Section 8.4. Before we start, we comment upon assumptions from standard black hole puzzles that we adopt, and point out critical subtleties therein that we will discuss later in Section 8.4.

8.1.1 Critical note about black hole assumptions

In the remainder of this section and in Section 8.2, we provisionally adopt the standard assumptions used in black-hole cloning and firewall arguments. These include the existence of suitable black hole radiation subsystems, the possibility of decoding them in the required spacetime region, an approximate inside/outside factorization, persistent interior records, and the identification of late radiation with near-horizon modes. These assumptions are not derived here and are known to be conceptually delicate in quantum gravity. The no-go results below should therefore be read as conditional results about the strengthened consequences of these assumptions. In Section 8.4 we return to these assumptions and highlight key issues that cast doubt upon the validity of these assumptions.

A finite-dimensional quantum-information model of the operationally relevant subsystems is assumed, with qubits and finite memories representing effective or code-subspace degrees of freedom. The no-go results are therefore conditional on the possibility of identifying, decoding, and coherently manipulating such effective subsystems, including subsystems encoded in the Hawking radiation. From the viewpoint of an observer outside the black hole, the black hole dynamics is modeled, for the purposes of the protocol, as implementing a unitary scrambling/encoding map U_R on the black hole, infalling systems and relevant environmental degrees of freedom, à la Hayden–Preskill [48]. It is further assumed that an appropriate recovery operation on the accessible Hawking radiation (possibly together with previously collected radiation or auxiliary systems) reconstructs the relevant code-subspace state of the laboratory, after which the inverse of the laboratory measurement can be applied as in the Wigner's friend scenario. Where (and when) and how such information recovery would occur is not prescribed, and these questions constitute an important point of scrutiny, as noted above and further discussed in Section 8.4.

It is further assumed that Alice and Bob hold systems that are sufficiently isolated within their lab and can be effectively modeled as qubits, upon which they will perform measurements. (This last assumption concerns only the relevant effective degrees of freedom, and does not require their gravitational fields to be perfectly shielded from their surroundings, for instance.)

8.1.2 Protocol and no-go theorem

Alice and Bob, residing in sealed labs, each measure their part of a Hardy state[1] [271]

$$|\psi_0\rangle_{S_AS_B} = \frac{1}{\sqrt{3}}(|00\rangle + |10\rangle + |11\rangle)_{S_AS_B} \tag{8.1}$$

in the computational basis with outcomes $a, b \in \{0, 1\}$, modeled unitarily as U_A, U_B, and next fall into an old black hole, i.e. a black hole past the Page time. Ursula retrieves and decodes the Hawking radiation in which the state of all of Alice's sealed lab is encoded (assuming an argument à la Hayden–Preskill [48], where for an outsider the black hole dynamics is assumed to be unitary and rapidly mixing), "undoes" Alice's measurement by applying the inverse unitary $U_A^\dagger$ and measures S_A in the $|\pm\rangle = {}^1/\sqrt{2}(|0\rangle \pm |1\rangle)$ basis with outcome $u \in \{\pm\}$.[2] Wigner acts analogously on Bob's Hawking radiation, obtaining $w \in \{\pm\}$. Alice and Bob, once inside the black hole, try to send their outcomes a, b towards Ursula and Wigner who fall into the black hole after having performed their measurements.[3]

Upon obtaining $u = -, w = -$, which has a nonzero probability of occurring, Ursula and Wigner reason as follows:

- $u = -$ implies $b = 1$ as $\langle -|_{S_A}\langle 0|_{S_B} |\psi_0\rangle_{S_AS_B} = 0$,
- $w = -$ implies $a = 0$ as $\langle 1|_{S_A} \langle -|_{S_B} |\psi_0\rangle_{S_AS_B} = 0$,
- but $a = 0, b = 1$ cannot occur as $\langle 01|_{S_AS_B} |\psi_0\rangle_{S_AS_B} = 0$,

and thus a contradiction is obtained. This contradiction becomes operational if Alice and Bob, once inside the black hole, succeed in sending their outcomes and predictions to Ursula and Wigner. We could also have formulated the protocol by having Alice and Bob reason about the outcomes of Ursula and Wigner.

[1] We will sometimes omit the tensor product symbol for ease of notation, i.e. write $|00\rangle$ or $|0\rangle\,|0\rangle$ for $|0\rangle \otimes |0\rangle$.

[2] One could also allow Alice (and Bob) to perform their measurement inside the black hole, and in that case question whether Ursula would still have to apply $U_A^\dagger$. I thank Eric Cavalcanti for this comment. This protocol may be more subtle, as it also involves the question of how Alice's (and Bob's) measurement inside is modeled from the outside.

[3] Note that for Ursula and Wigner to obtain a message from Alice and Bob, Ursula and Wigner must decode the Hawking radiation at locations where they can still fall into the (evaporating) black hole, and thus cannot do this at late times near null or timelike future infinity. One can also derive a contradiction when Ursula and Wigner remain outside the black hole, which is the content of Theorem 7. We comment more upon this in Section 8.4.

To explain in some more detail how the probabilities above are obtained, let us focus on the first, Ursula's claim about Bob's outcome b. We start with the initial Hardy state $|\psi_0\rangle$ of Eq. (8.1) for the systems $S_A \otimes S_B$, and Alice's and Bob's memories initialized in some states, the $|0\rangle$ states for example,

$$|\Psi_0\rangle := |\psi_0\rangle_{S_A S_B} |0\rangle_A |0\rangle_B . \tag{8.2}$$

Alice and Bob perform their measurements, which are modeled unitarily as U_A and U_B acting on Eq. (8.2), in this case CNOT gates with controls on S_A and S_B, as explained in Section 2.3, obtaining $U_A U_B |\Psi_0\rangle$.

We next make explicit the idealized recovery assumption used in the protocol. Let $L = S_A A S_B B$ denote the relevant joint laboratory degrees of freedom of Alice and Bob, let B_0 denote the black hole before the relevant scrambling and evaporation, let E denote previously emitted radiation, let R denote the subsequently emitted radiation accessible to Ursula and Wigner, and let B' denote the remaining black hole. From the outsider's point of view (Ursula's and Wigner's point of view), Alice's and Bob's labs get scrambled by the black hole into Hawking radiation by evaporation, implemented by a unitary as

$$V_{\mathrm{BH}} : \mathcal{H}_L \otimes \mathcal{H}_{B_0} \longrightarrow \mathcal{H}_R \otimes \mathcal{H}_{B'}. \tag{8.3}$$

We assume that, on the relevant laboratory code subspace L, there exists a coherent decoding isometry

$$W_{\mathrm{dec}} : \mathcal{H}_R \otimes \mathcal{H}_E \longrightarrow \mathcal{H}_{\widetilde{L}} \otimes \mathcal{H}_J \tag{8.4}$$

that decodes the relevant information from the radiation RE, i.e.

$$(W_{\mathrm{dec}} \otimes \mathbb{1}_{B'})(V_{\mathrm{BH}} \otimes \mathbb{1}_E) |\phi\rangle_L |\Omega\rangle_{B_0 E} = |\phi\rangle_{\widetilde{L}} |\chi\rangle_{JB'} \tag{8.5}$$

for every state $|\phi\rangle_L$ in that code subspace L. Here, $\widetilde{L}$ denotes the reconstructed laboratory, J denotes unused output degrees of freedom of the decoder, and the state $|\chi\rangle_{JB'}$ is independent of $|\phi\rangle_L$. We henceforth identify the reconstructed laboratory degrees of freedom $\widetilde{L}$ with the corresponding original labels L.

We consider now the case where Bob obtains $b = 0$, giving the state

$$\begin{aligned} &|0\rangle_B \langle 0|_B U_A U_B |\Psi_0\rangle \\ &= |00\rangle_{S_B B} \langle 0|_{S_B} U_A |\psi_0\rangle_{S_A S_B} |0\rangle_A \end{aligned} \tag{8.6}$$

up to normalization, where we used that

$$|0\rangle_B \langle 0|_B U_B |0\rangle_B = |00\rangle_{S_B B} \langle 0|_{S_B} . \tag{8.7}$$

Next, Alice's and Bob's labs get scrambled by V_{BH}, and Ursula and Wigner apply the decoding map W_{dec} to the radiation (which includes both the early and late radiation E and R), and thus obtain Eq. (8.6) after decoding by Eq. (8.5). Ursula next undoes Alice's measurement by applying $U_A^\dagger$, performs her measurement, where we consider the case where her outcome is $u = -$, giving

$$\begin{aligned} &|00\rangle_{S_B B} \langle -|_{S_A} U_A^\dagger \langle 0|_{S_B} U_A |\psi_0\rangle_{S_A S_B} |0\rangle_A \\ &= |00\rangle_{S_B B} \langle -|_{S_A} \langle 0|_{S_B} |\psi_0\rangle_{S_A S_B} |0\rangle_A \\ &= 0, \end{aligned} \tag{8.8}$$

using that $U_A^\dagger U_A = \mathbb{1}$ and $\langle -|_{S_A} \langle 0|_{S_B} |\psi_0\rangle_{S_A S_B} = 0$, resulting in a probability 0, $p(u = -, b = 0) = 0$. The other probabilities used in Ursula's and Wigner's reasoning are obtained analogously.

We have thus proved the following no-go theorem.

Theorem 6. *If a superobserver can decode the Hawking radiation and reconstruct the quantum state of an infalling observer, and, when falling in, still receive a short message from that observer, sent inside the black hole and assumed to be unaltered by actions outside, an operational contradiction is found with quantum theory's predictions.*

Compared to usual EWF arguments, here all of the outcomes a, b, u, w seem available on a single spatial slice, as can be seen on Fig. 8.1(b), and Theorem 6 does not require the assumption of absolute outcomes of performed measurements: it involves only empirical correlations accessible to a single observer.

If, as suggested in literature on the black hole cloning paradox (see Section 8.1.4), an observer catching the Hawking radiation from an infalling observer A cannot obtain any message inside from A due to time constraints, and so Ursula and Wigner are unable to obtain the outcomes of Alice and Bob inside, then we can still use our protocol to derive another no-go theorem.

Theorem 7. *If a superobserver can decode the Hawking radiation and reconstruct the quantum state of an infalling observer, a contradiction is found if (Ci) the outcome records of an observer inside the black hole remain coherent and untouched by operations outside, (Cii) observers can make valid Born rule predictions about measurement outcomes (without actually obtaining those outcomes, but where those outcome records remain*

untouched) and (Ciii) observers who can communicate and meet have to agree on their predictions.

8.1.3 Do the outcome records of observers inside get altered by operations outside?

In the protocol above, the correlation $p(u, w)$ can be obtained experimentally by Ursula and Wigner, who remain outside the black hole, and the correlation $p(a, b)$ by Alice and Bob who fall into the black hole together. However, if Alice and Bob fail to send a message with their outcomes inside the black hole to Ursula and Wigner, the correlations $p(u, b), p(a, w)$ are not verified experimentally as no single agent obtains both u, b, for example, which led to Theorem 7. So, in this case, Ursula and Wigner could just as well remain outside the black hole. But why does Ursula argue then that the joint correlations $p(u, b)$ of her prediction are those she would obtain if Bob had remained outside the black hole to compare her outcome to b? (Similarly for Wigner.)

Similarly to the Local Friendliness no-go theorem [12], one can try to verify this prediction in certain rounds of the protocol[4] (when Ursula and Wigner do not fall into the black hole), by allowing Alice to make a choice $x = 0, 1$ whether to leave her outcome outside the black hole or not (similarly for Bob with a choice y). In rounds where $x = 0$ she leaves a copy of her outcome a_O outside the black hole and the empirical correlations $p(a_O, w)$ are obtained outside the black hole. If Wigner can also access Alice's outcome a inside, he finds that $p(a_O, w, a|x = 0, y = 1) = \delta_{a_O,a} p(a_O, w|x = 0, y = 1)$, confirming consistency between inside and outside outcomes. Other choices of x, y test other statements. Implementing this idea in our first protocol we arrive at a no-go theorem where an assumption of no-signalling is used to argue that all these statements remain true regardless of the choices x, y.

An alternative way of motivating the validity of the quantum correlations $p(a, w)$, $p(u, b)$ is to note that if Bob fell into the black hole much earlier than Alice, Wigner could decode Bob's Hawking radiation and perform his measurement with outcome w, while Alice remains outside still, so that they can verify the correlation $p(a, w)$. An analogous argument holds for Alice falling in earlier and the verification of $p(u, b)$.

The arguments above suggest that the correlations $p(a, w), p(u, b), p(a, b), p(u, w)$ are indeed those predicted by quantum theory as explained in Eq. (8.8), if one can assign

[4] In these rounds of the protocol, in addition to the systems S_A, S_B, now also Alice and Bob and their lab environments are initialized in the same known initial states.

outcome values to the measurements of Alice and Bob who fell into the black hole. A possibility is that Ursula's and Wigner's actions erase Alice's and Bob's physical outcome records as in Wigner's friend scenarios. However, semiclassical causality arguments would argue that Ursula's and Wigner's actions outside the black hole cannot alter the experiences of Alice and Bob inside. Nevertheless, refuting this assumption, referred to as (Ci) in Theorems 6 and 7 and also present in the cloning paradox, still seems one of the most reasonable ways out if one believes in unitarity, compared to refuting assumptions (Cii) or $(Ciii)$ of Theorem 7. As we will discuss in Section 8.4.1, subtleties regarding black hole puzzles for infalling observers can be argued to make this option more natural.

8.1.4 Comparison with the no-cloning argument

The cloning paradox analyzed by Susskind and Thorlacius [47] and Hayden and Preskill [48] is similar in spirit to our protocol, but does not directly provide a logical contradiction on the level of operational probability predictions. Our result rules out any logically consistent theory that can describe the viewpoints of the different observers if they can use quantum theory for their own predictions, assuming memories inside the black hole remain unaltered by actions on the outside radiation.

In the cloning paradox, Alice falls into the black hole while carrying a quantum system A. Bob catches the Hawking radiation required to reconstruct A's state, and dives into the black hole. Alice, once inside the black hole, sends her system A (or quantum information about it) to Bob, who would observe a violation of quantum theory's no-cloning. Hayden and Preskill [48] suggested that Bob, who collects the Hawking radiation outside to reconstruct a system A's state, can just never retrieve the system A inside. Therefore, an operational violation of no-cloning seems protected. Namely, for Alice to send her qubit A to Bob, according to Hayden and Preskill, she must operate within a Planck time scale on her proper time to send information about her qubit to Bob for it to reach him before he hits the singularity[5]. But this means that Alice must send her information in super-Planckian frequencies, in a strong gravity regime, likely much different from low-energy gravity. The same argument against an actual operational contradiction holds for the protocol of Section 8.1.2. Nevertheless, the existence of these two copies in the cloning paradox remains problematic if one believes both views should be described by one single quantum theory. Similarly, in the protocol above, even if

[5]This is assuming the causal structure of the Schwarzschild black hole, and need not be the case for the classical causal structure of the rotating and charged black holes. However, the actual (quantum) interior of the latter is likely quite different from the classical one [489–503].

Ursula and Wigner can never get any messages from Alice and Bob inside the black hole, a puzzle remains, as stated in Theorem 7. In analogy with Theorem 7, we could phrase the no-cloning theorem as:

> If an observer can reconstruct a system S's quantum state from the Hawking radiation of an infalling quantum system, a violation of no-cloning is found if (Ci) the state of the quantum system S that fell into the black hole remains coherent and unaltered by operations outside, and (iv) quantum theory can describe the physics of the whole spacetime, including both the inside and outside of the black hole.

One could propose to refute (iv) here, and instead postulate a theory that describes the evolution of the whole spacetime. Our protocol and no-go theorem above show that this is impossible if no single observer can falsify quantum theory's operational predictions. We discuss these paradoxes further in Section 8.4.

8.1.5 Combining other extended Wigner's friend paradoxes with the cloning paradox

From the scenario outlined in Section 8.1.2, it is clear that we can combine other extended Wigner's friend arguments with the black hole cloning argument to produce no-go theorems. For example, we could come up with a scenario that combines the Local Friendliness no-go theorem with the black hole cloning paradox by having Ursula and Wigner choose x, y whether to reveal the outcome of Alice and Bob when they have decoded their Hawking radiation. Similarly, we could combine the Pusey–Barrett–Rudolph Wigner's friend argument of Chapter 5 with the black hole cloning paradox by having Wigner perform the PBR measurement on the decoded Hawking radiation from two friends Alice and Bob, who made some choices earlier while falling into the black hole.

8.2 Combining extended Wigner's friend paradoxes and the firewall argument

In this section we first recap the firewall paradox, after which we construct a protocol and no-go theorem that combines the firewall paradox and extended Wigner's friend arguments. Implications are discussed in Section 8.4.

8.2.1 Preliminaries: the firewall argument

The firewall argument. We follow the presentation of Ref. [107]. An infalling referee Bob holds a (distilled) qubit Q_B right outside an old black hole (i.e. past the Page time), being part of the late-time Hawking radiation. Two observers, Alice and Charlie, starting outside and inside the black hole, respectively, collaborate to provide two predictions $(P, \overline{P})$ to the referee Bob, for when he measures his qubit Q_B in the Z- and X-bases. Alice and Charlie win if the referee's test succeeds, i.e. if the referee Bob is given the correct predictions; this is called the *complementarity game* [107]. Charlie, inside the black hole, can distill a qubit Q_C (approximately) maximally entangled with Q_B by the assumption of high entanglement across the horizon [392, 504] (assuming semiclassical gravity), and Alice, holding the early-time Hawking radiation outside the black hole, can distill a qubit Q_A (approximately) maximally entangled with Q_B by Page's argument [62, 63]. Alice and Charlie measure their qubits Q_A, Q_C in the X- and Z-bases, respectively, obtaining outcomes a, c. As Charlie's qubit is maximally entangled with Q_B, he predicts the referee's outcome b to be the same as his outcome c when he measures his qubit in the Z-basis, and similarly for Alice predicting outcome a for the referee when he measures in the X-basis. They send their predictions (a, c) to the referee who falls into the black hole (to retrieve Charlie's message and outcome). The predictions $P = a, \overline{P} = c$ yield a winning probability of 1, violating the quantum bound[6] of $\frac{1}{2} + \frac{1}{\sqrt{8}} < 1$ [107, 505] for winning the complementarity game. This violation is based on a violation of monogamy of entanglement.

We can also construct an extension of the firewall argument by giving Charlie and the referee Bob measurement choices x, y, resulting in a violation of no-signalling based on the violation of monogamy of entanglement (see Appendix C of Ref. [386]).

If some observer can obtain all outcomes of Alice, Bob and Charlie, an operational violation of quantum theory's formalism is found in the firewall argument. Such operational violation is more plausible than in the black hole cloning paradox, as no observer needs to wait for the Hawking radiation of an infalling observer to have come out. In fact, in typical treatments of the firewall argument, Alice, Charlie and the referee Bob are played by a single observer; this is possible if Alice's distilling operation can be performed sufficiently rapidly so she can still obtain Charlie's mode inside the black hole. A potential problem in both of these paradoxes lies in the complexity of Alice's operation of distilling

[6] The 'quantum bound' means the maximal probability of winning the complementarity game achievable by any strategy permitted by quantum theory.

the qubit Q_A from the early radiation, which could require a time exponentially long in the black hole entropy [49, 53, 475] compared to its polynomial evaporation time. However, with exponentially long preprocessing time such long decoding time could be avoided, as suggested in a toy black hole representation in Ref. [480].

The firewall paradox as a no-go theorem. The firewall argument can be stated as a no-go theorem in the following way.

Theorem 8 (Firewall paradox)**.** *The firewall argument shows that the following four assumptions are incompatible:*

(Fi) the early radiation is maximally entangled with the late radiation—motivated by unitarity and the Page curve—and (Fii) an observer can distill a qubit Q_A of the early radiation maximally entangled with a qubit mode Q_B of the late radiation that is just outside the horizon;

(Fiii) the late radiation modes just outside the horizon are highly entangled with modes inside the black hole such that inside one can distill a qubit Q_C maximally entangled with Q_B—motivated by the equivalence principle and high entanglement across the horizon as in QFT on curved spacetime;

(iv) quantum theory can describe the whole evolution comprising both the inside and outside of the black hole (including all radiation).

One way to evade this no-go theorem would be to alter assumption (iv)[7], by considering a more general theory than quantum theory that describes the evolution of both the inside and outside of the black hole. This route will not work if one accepts the existence of superobservers along with the persistence of outcome records of the black hole interior as in (Ci): any theory that would do so would find an experimental violation of quantum theory's predictions. We prove this claim in the next section by presenting a protocol that combines Wigner's friend with the firewall paradox. Alternatively, one can introduce measurement choices x, y for Charlie and the referee Bob in the firewall paradox, and use the monogamy of entanglement violation to derive a contradiction of $(Fi) - (Fiii)$ with no-signalling (as defined in the semiclassical causal structure) as we prove in Appendix A.2.

[7] In typical treatments of the firewall and cloning paradoxes, assumption (iv) is not explicitly stated, but implicitly assumed.

8.2.2 Protocol and no-go theorem

Consider the following protocol, laid out stepwise and sketched in Fig. 8.2, which can be seen as combining Wigner's friend arguments and the firewall paradox.

Initial set-up. The argument involves four observers Alice, Bob[8], Ursula and Wigner, who all start outside the black hole. Similarly as in the firewall paradox, Bob holds a qubit mode of the late radiation close to an old black hole, which he gathered in free fall. Alice holds the early radiation, and decodes a qubit S_A maximally entangled with Bob's qubit S_B, the same starting set-up as in the firewall paradox, such that the systems they hold are in the Bell state $(|00\rangle + |11\rangle)_{S_A S_B}/\sqrt{2}$. We also assume that Alice and Bob are each located in a sealed lab. The measurements that will be performed on the systems S_A, S_B have correlations that violate the CHSH inequality [236], such that there exists no global distribution reproducing all of these correlations [34, 42, 247, 342].

Step 1. Hovering outside the black hole (having stopped his free fall), Bob measures his system S_B in the $|\pm\rangle$-basis, with outcome labelled with b, while Alice measures in the basis $|a_0\rangle, |a_1\rangle$ with outcome denoted $a = 0, 1$ and

$$|a_0\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle, \quad |a_1\rangle = -\sin\theta|0\rangle + \cos\theta|1\rangle, \tag{8.9}$$

with $\theta = \pi/8$. Next, Alice and Bob come together and compare their outcomes, verifying the correlation $p(a, b)$ (verifying the consistency of the joint outcomes a, b with quantum theory's prediction for $p(a, b)$). Ursula, having perfect control over everything that happens in Bob's lab, subsequently undoes this verification[9].

Step 2. Alice falls into the black hole and measures a qubit S_W that is maximally entangled with Bob's qubit S_B, i.e. whose joint state is given by $(|00\rangle + |11\rangle)_{S_W S_B}/\sqrt{2}$; this is the use of maximal entanglement across the horizon as in the firewall paradox.[10] Alice measures the qubit S_W in the computational $|0\rangle, |1\rangle$ basis, labeling her outcome by a_w. She then verifies the correlation $p(a, a_w)$.

Step 3. In the meantime, outside the black hole, Ursula decodes Alice's Hawking radiation, and undoes her measurement by restoring Alice's lab back to its initial state.

[8] Compared to the firewall paradox in the complementarity game presentation above in Section 8.2.1, Bob is not a referee in this protocol, just an agent performing a measurement.

[9] As Ursula has perfect quantum control over Bob's sealed lab, she can apply the inverse of the unitary verification operation by Alice and Bob.

[10] Here we neglect the potential issue of long decoding times of the Hawking radiation as mentioned in Section 8.2.1.

Next, she performs the measurement $|u_0\rangle, |u_1\rangle$ with outcome denoted u and

$$|u_0\rangle = \cos\theta|0\rangle - \sin\theta|1\rangle, \quad |u_1\rangle = \sin\theta|0\rangle + \cos\theta|1\rangle, \tag{8.10}$$

with $\theta = \pi/8$. Having obtained her outcome u, she asks Bob for his outcome and they verify the correlation $p(u, b)$. After this, Wigner undoes this verification and lets Bob fall into the black hole.

Step 4. Wigner decodes the Hawking radiation from Bob's lab outside the black hole, and undoes Bob's measurement to reset his lab back to its initial state. After this, Wigner measures Bob's qubit S_B in the computational $|0\rangle, |1\rangle$ basis, obtaining an outcome labelled by w. Ursula then asks for Wigner's outcome and they verify the correlation $p(u, w)$. Now, because S_B and S_W were maximally entangled as $(|00\rangle + |11\rangle)_{S_W S_B}/\sqrt{2}$, and w and a_w were obtained by measuring S_B and S_W in the computational basis, quantum theory tells us that $a_w = w$. We use this fact to argue that the correlation $p(a, a_w)$ would be the same as the correlation $p(a, w)$, which arises from outcomes a, w of measurements in the $|a_0\rangle, |a_1\rangle$ and computational basis, respectively, on the maximally entangled state $(|00\rangle + |11\rangle)_{S_A S_B}/\sqrt{2}$.

Deriving a contradiction. Thus, in the course of this scenario, all of the correlations $p(a, b), p(u, b), p(a, a_w), p(u, w)$ were verified, with moreover $a_w = w$. In addition, all of these outcomes a, b, u, a_w, w still exist, some inside the black hole and some outside. Thus, they seemingly coexist on a spatial slice still, suggesting that all of these correlations should come from a joint probability distribution $p(a, b, u, a_w, w)$. However, this cannot be true as the verified correlations are chosen to be of the sort that violate the CHSH inequality, such that no such global distribution $p(a, b, u, a_w, w)$ reproducing these correlations $p(a, b), p(u, b), p(a, a_w), p(u, w)$ with $a_w = w$ exists [34, 42, 236, 247, 342]. We thus find a contradiction.

This argument leads to the following no-go theorem.

Theorem 9 (EWF and firewalls)**.** *If a superobserver can decode the Hawking radiation and reconstruct the quantum state of an infalling observer, then under the assumptions (Fi)-(Fiii) of the firewall paradox of Theorem 8 and (Ci) of Theorem 7, a contradiction is found with quantum theory's predictions for joint correlations.*

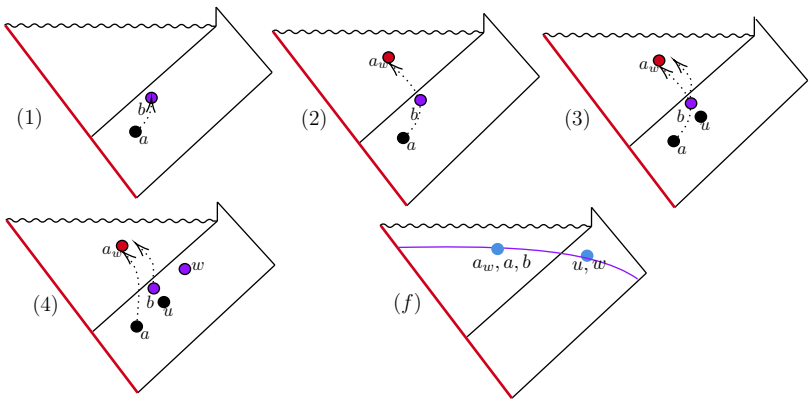

Figure 8.2: Stepwise explanation of the EWF-firewall protocol. (1) Alice and Bob hold a maximally entangled state, obtained from high entanglement between early and late radiation, perform their measurements with outcomes a, b and verify the correlation $p(a, b)$; this verification action is then undone by Ursula. (2) Alice falls into the black hole and measures a qubit S_W that is maximally entangled with Bob's qubit, obtained from high entanglement across the horizon, verifying the correlation $p(a, a_w)$, while Bob remains outside. (3) Ursula catches the Hawking radiation of Alice's lab, undoes her measurement and performs a measurement with outcome u so that she can verify the correlation $p(u, b)$ with Bob. Next, Wigner undoes this verification, and Bob falls into the black hole. (4) Wigner catches the Hawking radiation from Bob's lab, undoes Bob's measurement and performs a measurement with outcome w, verifying the correlation $p(u, w)$ with Ursula. Summarizing, in steps (1), (2), (3) and (4), the correlations $p(a, b), p(a, a_w), p(u, b)$ and $p(u, w)$ are verified. The final picture (f) shows how seemingly all outcomes a_w, a, b, u, w coexist, contradicting the fact that the verified correlations violate the CHSH inequality and cannot arise from a single, global probability distribution. We comment on these Penrose diagrams and locations of information recovery in Section 8.4.

8.3 Relating to the work by Hausmann and Renner

Hausmann and Renner (HR) [107] describe two Wigner's friend protocols and two protocols involving a black hole, and draw analogies between Wigner's friend and black hole paradoxes. We have extended their work by combining an extended Wigner's friend protocol [11–14, 19] with black hole physics into a unified argument. HR's first black hole argument is similar to the cloning paradox, and can be seen as combining a single Wigner's friend scenario with the cloning paradox. Their second no-go theorem, which we have described in Section 8.2.1, is a rephrasing of the firewall argument where the non-monogamy of entanglement is used to violate the quantum bound of the complementarity game.

How Hausmann and Renner relate to Wigner's friend. Theorem 6 assumes that infalling observers can still send a message to the observers catching their Hawking

radiation who fall into the black hole later, an assumption that may be problematic as argued in the context of the black hole cloning paradox (see Section 8.1.4), while Theorem 7 derives a no-go theorem without such assumption, based on the same protocol. In the firewall paradox, the possibility of one single observer bringing the problematic aspects inside and outside the black hole together is more plausible as no observer needs to wait outside the black hole to catch the Hawking radiation of an infalling observer before falling into the black hole. Hausmann and Renner [107] present a no-go theorem based on the firewall paradox as paraphrased in Section 8.2.1, but with assumptions phrased differently than the ones of Theorem 8. A key assumption in their no-go theorem is a consistency assumption ($\mathcal{C}$), which allows Alice, inside the black hole, to reason about Bob outside the black hole. We can paraphrase the consistency assumption ($\mathcal{C}$), also used in EWF arguments such as the FR paradox [11], as follows:

> ($\mathcal{C}$): If an observer B, modeled classically by observer A (i.e., above A's Heisenberg cut), knows that a statement s about some measurement outcome(s) is true from reasoning with a theory that A accepts, and A knows about this reasoning of B, then s is true for A as well.

Notably, assumption ($\mathcal{C}$) is already violated in certain classical paradoxes [120, 292], and is thus perhaps too strong. Therefore, one could apply the same criticism to Hausmann and Renner's theorems [107, Theorems 3 and 5]. However, the FR paradox can be slightly rephrased such that only a weaker consistency assumption is required [19], which we refer to as ($\mathcal{C}'$), where observers A and B only need to take each other's statements to be true if they can communicate and physically meet. We use a similar condition to ($\mathcal{C}'$) in Theorem 7, the combination of assumptions (Cii) and $(Ciii)$ there. In the EWF-firewall argument of Theorem 9, we do not need a condition for the consistency of reasoning or predictions such as ($\mathcal{C}'$), as all involved correlations are actually observed, similarly as in typical formulations of the firewall paradox.

In fact, in the firewall paradox formulation of Section 8.2.1 (per Hausmann and Renner [107]), the assumption ($\mathcal{C}'$) could be used instead of ($\mathcal{C}$) if Bob were to fall into the black hole to meet Alice. However, Hausmann and Renner refrain from letting Bob fall into the black hole, as they model his reasoning while he remains outside the black hole, used in the assumptions that as an outsider he can model the Hawking radiation process as a typical unitary for the early radiation and the modes at the stretched horizon (which will constitute the late radiation). Assuming no-retrocausality, one could argue that Bob only needs to model this unitary up to the point where he could verify his claims

by asking the referee holding the modes at the stretched horizon, to use $(\mathcal{C}')$ instead of $(\mathcal{C})$. We will not discuss this matter in more detail here.

The refined version of the FR paradox that uses $(\mathcal{C}')$ instead of $(\mathcal{C})$ [19], as well as other EWF arguments [11–13, 28], can be used to argue against absolute, single-valued outcomes of performed measurements, in favor of relationalism [19, 288, 300, 320, 321, 325], assuming the existence of superobservers. However, the no-go theorems in this chapter serve another purpose. Naive treatments of the interior and exterior of the black hole suggest that all relevant outcomes coexist on a single spatial slice, and thus arguing that these outcomes would not coexist as in EWF paradoxes such as the (refined) FR paradox seems somewhat odd here. Therefore, these no-go theorems rather suggest that these naive treatments of black hole physics must be false.

8.4 Discussion

We have presented two scenarios combining Wigner's friend and black hole physics into a unified argument, resulting in Theorems 6, 7 and 9 that extend the black hole cloning and firewall paradoxes. Specifically, the cloning paradox leaves room for a theory that describes the interior and exterior of the black hole to evade the no-cloning argument while not giving up on the standard black hole assumptions without violating quantum theory's predictions. With the scenarios in Section 8.1 we close this loophole, and show that no such theory exists if observers cannot witness an experimental violation of quantum theory's predictions. Similarly, the firewall paradox leaves room for such a theory evading the firewall argument while not giving up on the standard black hole assumptions going into the firewall paradox without violating quantum theory's predictions. In Section 8.2 we show that such theory must violate no-signalling, and gives a violation of quantum theory's predictions when assuming the persistence of outcome records in the black hole interior. Consequently, the problem must lie in other assumptions of the firewall and cloning paradoxes to prevent such violations.

In Chapter 7 we have argued that there are three answers to the black hole information puzzle, (i) information loss[11] [506], (ii) remnant-like objects holding the information, and

[11] A central argument for information loss is that, semiclassically, the entanglement of the Hawking radiation with degrees of freedom behind the horizon persists after evaporation, so that a hypersurface containing the radiation in the post-evaporation region is not a Cauchy surface and the state restricted to it is mixed [506]. Whether such argument survives quantum gravity and superposed spacetimes

(iii) aspects that were overlooked in Hawking's calculations, such as soft hair, superposed geometries, replica wormholes and islands, for example. It is within (iii) that paradoxes for infalling observers such as the cloning and firewall paradoxes take place. In the remainder of this section, we will highlight subtleties in these paradoxes, and point toward further directions. As commented upon in Chapter 7, one possibility is to invoke modifications at the level of the horizon, or its status, such as firewalls, fuzzballs and gravastars. We will instead focus on subtleties that are present without a priori invoking such new structures, and briefly discuss their implications.

8.4.1 Subtleties in black hole paradoxes

a. Conservation laws and superposed geometries. First, by conservation laws [110, 111], one expects an old black hole to be highly superposed, for example with a highly superposed mass [112], also advocated for in Refs. [70, 113–116]. Moreover, a black hole's soft hair in 4d asymptotically flat spacetime implies that the classical vacuum is highly degenerate [102, 118]. Therefore, drawing Penrose diagrams[12] with single-spacetime slices into the bulk of an old black hole may be problematic, and observers falling into the black hole can get entangled with its properties through the gravitational interaction, or even measure them.[13]

b. Locality and gauge invariance. Secondly, the notions of locality and gauge invariance[14] in (quantum) gravity are subtle, and may have substantive implications for black hole (infaller) paradoxes [100, 481, 512–516]. To make a localized observable gauge-invariant, one can use relational observables or gravitational dressings [65, 481, 512–514, 517–529], with nontrivial effects regarding spacetime locality. Such dressings have been argued to be of crucial importance for the mechanism behind islands [100]. A

remains unclear. For example, holography [64, 507, 508] and the potential failure of the split property in quantum gravity [98] would imply that the 'interior' degrees of freedom are not necessarily independent of those in the (post-evaporation) region where the Hawking radiation resides.

[12] Additionally, it is unclear how to continue a Penrose diagram post-evaporation as the near-singularity region is one of strong quantum gravity, see for example Fig. 8.3 for two different proposals. This is why in later Penrose diagrams of Fig. 8.4 we will add a question mark '?' for how exactly the post-evaporation spacetime should be pictured.

[13] One possibility would be to construct an effective spacetime metric from the correlators in the superposed spacetimes [119, 509, 510]; this requires further exploration and may not necessarily yield the standard black hole metric. See also Ref. [511] for semiclassical effective metrics as averages over superposed metrics.

[14] I thank David Mattingly for many interesting discussions and exchanges on these issues.

related problem is the impossibility to locally factorize the Hilbert spaces of the outside radiation and inside of the black hole in the bulk in quantum gravity [98, 512, 513].

c. Where is a Page curve obtained? At future null infinity? Related to the previous point, one could argue that unitarity from the radiation and the Page curve[15] are only obtained near or at future (null) infinity in asymptotically flat spacetimes, which is also where the radiation was collected in an asymptotically flat study of islands [79], and that gauge-invariantly identifying the near horizon modes and their evolution to form the late radiation arriving at future (null) infinity has nontrivial implications. Closer to an old black hole, one can still have the ability to measure or get entangled with the (superposed) gravitational field, isolating the radiation and decoding information from it would involve a nontrivial construction in the bulk, with the backreaction of that construction being potentially non-negligible [550], and one has the failure of factorization per b there; see also Fig. 8.4.

Therefore, regarding the cloning paradox, even if the Page curve suggests that one would only require a few qubits from an old black hole to reconstruct the quantum state of a system S that just fell into the black hole [48], one might need to recover those qubits at or near future (null) infinity, with no 'nice' slices [49, 473, 525, 551–554] hugging future null infinity until late times in an evaporating black hole spacetime while running into the black hole interior containing the system S that fell in, as in Fig. 8.4. Additionally, potential issues of nonlocal and non-adiabatic effects on nice slices have been expressed before [473, 525, 553].

If information recovery and a Page curve are only to be expected at or near future (null) infinity I^+, a possibility could be that operations on the radiation do change the memory of an infalling observer, but, with the same causality of Wigner's friend; only when that memory arrives at future (null) infinity encoded in the radiation. Such causality picture, with experiences during infall taking place earlier than the collection of that information at infinity, was advocated for recently in Ref. [488]. Naturally, questions of the full experience of an infalling observer are still limited by the black hole singularity, and dependent upon its potential resolution. Speculations aside, the issues highlighted in

[15] If gravity is holographic [64, 74, 98, 481, 507, 512, 515, 530–549], one might have to couple the boundary, such as the AdS one in recent replica-trick calculations [68, 75, 77, 391, 396, 397], to a non-gravitational bath to store the radiation and obtain a Page curve, argued to lead to massive gravitons through a Higgs mechanism in Refs. [65, 95, 96, 98–100]. Another possibility is to couple the black hole to a gravitational bath, and define the black hole and bath system (the radiation) through a non-geometric split [66–68].

this section call for an operational view on causality in quantum gravity [282, 555–558].

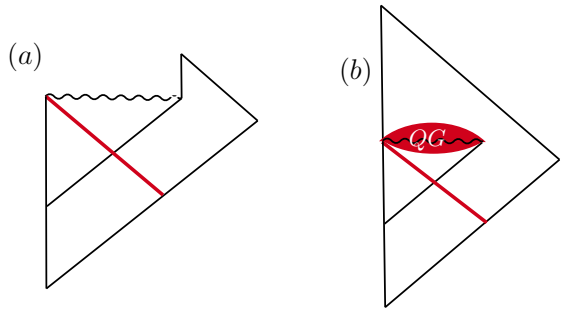

Figure 8.3: To discuss the evaporation of black holes and unitarity, the post-evaporation spacetime must be pictured as well. Two possibilities are (a): continuing past the singularity, and (b) the same continuation but for a nonsingular black hole; see also Refs. [77, 101, 394, 402–404, 559–566].

d. More subtleties and observer-dependencies in the firewall paradox. Fourth, as for the firewall paradox, the appearance of superposed geometries for old black holes was already mentioned above in a, conflicting with the assumption $(Fiii)$ in Theorem 8 of high entanglement across the horizon derived from QFT on a single spacetime, as well as the non-factorization regarding locality and gauge invariance per b for specifying the early radiation and near horizon (interior and exterior) modes.

Moreover, also in QFT, the experience of an infalling observer, and the entanglement they see, may be very different from an observer remaining outside (for longer) [567, 568]. Consequently, the descriptions of the black hole's evolution and entanglement profile for different observers could be quite distinct [569, 570], with entanglement known to be a frame-dependent notion [329] and potential subtleties regarding the equivalence principle for infalling observers [567, 568, 571–573]. As mentioned in the previous point c, assumption (Fii) of the firewall paradox as presented in Theorem 8 can also be questioned, with late-time radiation modes only required to purify the early radiation when arriving at infinity, and need not necessarily be trivially identified with near-horizon modes.

e. Infaller's backreaction and initial states. Fifth, the backreaction of an infalling object or observer is a nontrivial effect [73, 75, 116, 394, 478, 574–583]. Backreaction due to infalling objects and aspects of ultrarelativistic scattering have been investigated using shockwaves [574, 575, 577–579, 582–585], but can also be probed using (black hole) eikonal approximations to scattering [576, 581, 586–598].

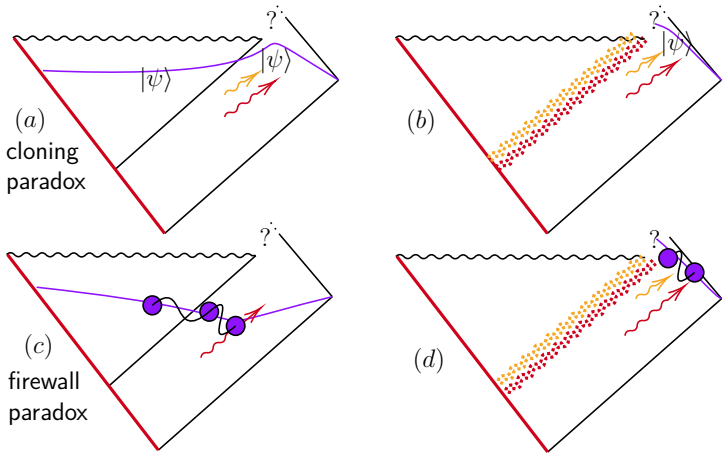

Figure 8.4: A critique of the cloning paradox (a), with the state $|\psi\rangle$ cloned on a spatial surface, and the firewall paradox (c), with three systems violating monogamy of entanglement, the inside mode highly entangled with the mode just outside the horizon and the early radiation also highly entangled with the same late mode just outside the horizon. Diagrams (b) and (d) give an alternative views for the cloning and firewall paradox, respectively. Namely, in order to recover the information, it is only at or near I^+ that one must be able to reconstruct the state $|\psi\rangle$ of an infalling system from the radiation and potentially remaining gravitational degrees of freedom. Similarly, it is only at or near I^+ that the early radiation must be highly entangled with the late radiation. Furthermore, the experiences of infalling observers might be quite different as they get entangled with the superposed, backreacted geometry (in which a notion of single spatial slices may be questionable as well).

Furthermore, maintaining the coherence of the quantum information encoded in an infalling system as observed by a co-infalling observer, a crucial assumption in the cloning paradox for example, requires careful shielding of that quantum information to prevent leakage and decoherence. This is complicated by the standard assumption that gravity does not allow for shielding [599–601], implying that such quantum information should be encoded in other degrees of freedom. Recent arguments of decoherence of superpositions due to (soft) radiation leaking into horizons or null infinity demonstrate such complications [602–605].

Relatedly, a full treatment of the unitarity problem of black hole evolution requires specification of the initial *quantum* state for black hole formation, where literature separates into collapsing shells [409, 452, 606–617], high-energy collisions of two particles [587–591, 593, 597, 618–635], holographic investigations of collapsing shells [636] and colliding particles [637], and Euclidean path integral preparations of black holes in two-dimensional gravity [395, 638].

8.4.2 Implications and future directions

The considerations above raise several doubts upon typical assumptions in black hole paradoxes for infalling observers, which also underlie our no-go theorems. Specifically, the locality and gauge-invariance issues of b, further complicate the superposed spacetime of an old black hole per a, and the related problems of causality, and how and where information is recovered from the radiation per c, as well as the infaller's backreaction in e, question how independent the relevant interior and exterior operators in black hole paradoxes are. These elements thus question the validity of assumption (Ci) in the cloning paradox and Theorems 6 and 7, that quantum information 'inside' remains unaltered by operations 'outside' the black hole, with inside and outside referring to a semiclassical picture.

Furthermore, superposed geometries and soft hair, non-factorization, the location of information retrieval and observer-dependencies as in a, b, c and d present challenges for assumptions (Fii) and $(Fiii)$ of the firewall paradox and Theorem 9, namely the relevance of high entanglement across the horizon and the identification of late-time (purifying) radiation modes with near horizon modes.

Therefore, with the aim of progressing towards a better physical understanding, operational investigations of assumptions and explicit modeling of infalling quantum systems in the context of black hole puzzles are an interesting avenue, with potential for lab experiments in analogue gravity [639–648]. One such possibility would entail the use of quantum detectors, where theoretical studies with detectors in (nonclassical) spacetimes have given us a rich plethora of results already [119, 567, 568, 572, 573, 649–681]. Furthermore, as black hole puzzles concern different observers' perspectives, questions of gauge invariance, dressings, locality and soft and edge modes, with moreover radiation leaking into or across (reference frames at) null infinity, the paradigm of quantum reference frames [151, 329, 465, 466, 469–471, 526, 682–685] may provide a useful framework for further investigations of some of these subtleties. We move towards such investigations in the next chapter, combining elements of quantum reference frames with detectors in a nonclassical gravitational setting.

Finally, computational complexity and quantum error correction have been associated with black holes, holography and the decoding of Hawking radiation on numerous occasions [49, 53, 474–477, 479, 480, 686–696], but have not been probed in the context of Wigner's friend. Where are *our* superobservers? Why can we trust our memories, for

example for formulating physical laws? If Wigner can manipulate the brain of his friend, Wigner could implant memories and experiences that would lead the friend to different physical laws, or even nonsensical results. We can imagine a similar situation when an observer, let's say a bee, falls into a fire. It burns to ashes, but we expect that when carefully keeping track of all relevant pieces of information, we can reconstruct the bee in the (classical) state right before it fell in. Now if we could perform such a complex operation, then perhaps we could also manipulate its brain into thinking that right before it fell into the fire the world was filled with flying crocodiles. Could it be that emergence and computational complexity are important concepts regarding consistency, Wigner's friend experiments and related (self-referential) puzzles such as Boltzmann brains [292, 382]?

The apparent absence of superobservers [697] is implied by the emergence of (kinematic and dynamical) classicality [5, 6, 307, 698–708], which have been linked to classical chaos and random matrices [5, 709, 710]. Moreover, topics such as computational complexity, chaos, quantum error correction and scrambling appear in measurement-induced phase transitions [711–715]. It would be interesting to further probe connections between these areas, and review the concept of an observer in this light.

— 9 —

Particle detector near position-superposed black hole

In the previous chapter, we have outlined several subtleties in standard assumptions of black hole paradoxes. Among the methods that can be employed to further probe these assumptions are quantum reference frames (QRFs) and detectors. In this chapter, we present an analysis combining these elements, by calculating the response of an Unruh–DeWitt detector in a 2+1d spacetime that contains a BTZ black hole in a superposition of locations as viewed from a reference frame adapted to the detector. We do so without relying on the details of a full theory of quantum gravity, but showing that this physical situation can be equivalently (and more conveniently) described in a different reference frame as a detector in a superposition of different locations in a fully classical gravitational field. In this frame, we can calculate the detector's evolution without involving any quantum property of gravity. Next, we perform a second QRF transformation to return to the original frame, where the black hole is in superposition, and examine the resulting dynamics. Our findings demonstrate that a joint measurement of the detector and the black hole's position can reveal effects of the superposition, manifesting as interference fringes. We define a measurement whose outcome probabilities contain a nonclassical contribution that would be absent for a black hole described by a classical mixture of positions.

Our methods and setup are inspired by Ref. [119], which studies a BTZ black hole in superposition of different masses. By contrast, we place the black hole in superposition of locations (as viewed from a reference frame adapted to the detector). Crucially, we find similar but different results: the resulting interference pattern is smooth - showing no sharp peaks—unlike the pronounced peaks presented in Ref. [119]. This further supports

the interpretation in Ref. [119] that such peaks are a consequence of black hole mass quantization [716–720]. We show analytically how this difference arises from singularities in the spectrum probed by the detector.

The rest of this chapter is organized as follows. In Section 9.1, we further motivate our approach, combining bottom-up quantum gravity, quantum reference frames and detectors. Next, in Section 9.2 we introduce the necessary concepts for our calculations: we provide a short pedagogical introduction to the BTZ black hole, Unruh–DeWitt detectors and quantum reference frames. The reader who is already familiar with these concepts can directly read Sections 9.3 and 9.4, where we describe the physical scenario we consider, and present our theoretical, numerical results and corresponding analytical expressions. Finally, we conclude in Section 9.5. We work in natural units, i.e. $\hbar = k_B = c = 8G = 1$.

9.1 Motivation: bottom-up quantum gravity, quantum reference frames and detectors

Black holes serve as a prime example of a natural setting where quantum theory intersects with general relativity. Ideally, a full theory of quantum gravity is needed to completely describe the behaviour of black holes as specific cases within a broader framework. This approach, often referred to as "top-down", has so far struggled to yield concrete predictions due to the challenges in making calculations with existing models.

Crucially, one of the most intriguing aspects of black holes——their emission of Hawking radiation—was predicted using a different, "bottom-up" approach. Without relying on a full theory of quantum gravity, but instead using only established principles from general relativity and quantum theory, Hawking predicted that black holes emit (approximately) thermal radiation [45]. This discovery led to fundamental puzzles in theoretical physics, such as the black hole information puzzle [46, 49, 388], and the cloning [48, 72] and firewall paradoxes [50–52, 62, 389], which we discussed in Chapters 7 and 8.

Other examples of "bottom-up" approaches include theoretical proposals to study quantum-gravitational effects that would be absent if gravity was classical, such as table-top quantum gravity experiments [721–729], or the study of quantum superpositions of semiclassical spacetimes [119, 664, 666, 730, 731]. In the last framework, quantum

reference frames (QRFs) [328, 329, 464, 469–472, 526, 683, 732–738] have caught recent attention, as a tool to study quantum gravitational phenomena: when spacetime exhibits non-classical behaviour, classical coordinates may no longer provide an adequate description of the physical situation, necessitating the use of a quantum framework for reference frames. Moreover, QRFs have also been linked to asymptotic symmetries [110, 438–440], memory [433–437], and soft and edge modes [739–742] in Refs. [469, 470, 743, 744], which endow asymptotically flat spacetimes with soft hair through infrared physics [102], argued to play an important role for the black hole information puzzle, for example [101, 103, 105, 745].

An operational way to analyze the behaviour of black holes and Hawking radiation is by probing the radiation with a particle detector, such as the Unruh–DeWitt detector [197]. This approach has recently regained much attention: work has been done on the response of detectors around black holes [490, 567, 649–658, 746] and in AdS spacetime [747], infalling detectors crossing the black hole horizon [568, 572, 573, 659, 660], detectors in a superposed trajectory [661–663], detectors in superposed quotiented Minkowski [664] and Rindler spacetimes [665], in superposed BTZ black holes [119, 666], superposed detectors in de Sitter and Schwarzschild spacetime [655, 667], entanglement harvesting [649, 668–677, 748–751], acceleration- and gravity-induced transparency effects [678, 679], anti-Unruh [752, 753] and anti-Hawking [656, 754] phenomena, covariantly smeared detectors in spacetime [755], etc. Furthermore, more complex models for detectors have been proposed in [756–759], used for example for investigating the Unruh effect [758, 760].

9.2 Introducing BTZ black holes, Unruh–DeWitt detectors and quantum reference frames

9.2.1 BTZ Black hole

The BTZ black hole [109, 761, 762] is a 2+1-dimensional black hole solution to Einstein's equation with negative cosmological constant $\Lambda = -1/l^2$, asymptotically anti-de Sitter (AdS_3) with radius l. It is particularly interesting as its field correlation functions have analytically closed forms [109, 119, 649, 652, 653, 746, 763], making theoretical explorations more tractable.

We consider here a non-rotating and uncharged BTZ black hole, with metric

$$\mathrm{d}s^2 = -f(r)\,\mathrm{d}t^2 + f(r)^{-1}\,\mathrm{d}r^2 + r^2\,\mathrm{d}\varphi^2, \tag{9.1}$$

with $f(r) = r^2/l^2 - M$ and coordinate range $\sqrt{M}l < r < \infty, -\infty < t < \infty, \varphi \in [0, 2\pi)$.

The black hole has an event horizon at $r_h = \sqrt{M}l$, with maximally extended Penrose diagram as in Fig. 9.1. Notably, being a 2+1-dimensional gravitational spacetime, the BTZ black hole has no Newtonian limit [109, 764–766], but other limits such as the Newton–Hooke one can be considered [767–770].

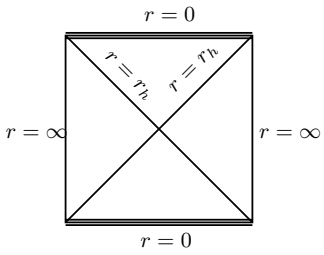

Figure 9.1: The (maximally) extended Penrose diagram of the BTZ black hole [771].

Spatial infinity is timelike, so that information can enter and leave at spatial infinity. Therefore, boundary conditions must be imposed for a field theory on the BTZ spacetime. In general, 2+1d gravity has no propagating degrees of freedom, but it is due to this boundary that nontrivial dynamics can occur, and so-called 'boundary gravitons' can appear in a quantum theory of gravity [762, 763, 771]. The BTZ black hole can be obtained from the universal covering space of AdS_3 [772] by means of the quotient under $\Gamma_M : \varphi \to \varphi + 2\pi\sqrt{M}$, see Section B.1. This fact will allow for correlation functions of quantum fields in BTZ black hole spacetime to be calculated precisely using correlations in AdS_3. The temperature assigned[1] to a BTZ black hole equals $T_0 = \sqrt{M}/(2\pi l)$ [109], and indeed a static detector around a BTZ black hole detects a red- or blueshifted local Tolman temperature [774, 775] $T_0/\sqrt{-g_{00}}$ [653, 654].

A test particle at fixed radius R in space moving only in time has proper time that relates to coordinate time t as $g_{00}(dt/d\tau)^2 = -1$ so

$$\tau = \sqrt{-g_{00}}t = \sqrt{f(R)}t =: \gamma(R)t = \sqrt{\frac{R^2}{l^2} - M}\,t, \tag{9.2}$$

[1] By a 'temperature assigned to a black hole', we mean the mutually consistent 'temperature' parameter, identified through the thermal spectrum of exterior quantum fields and, in equilibrium, as the KMS temperature of the exterior Hartle–Hawking state [45, 193, 746], the periodicity of Euclidean time required for regularity at the horizon [203, 761], or from thermodynamic considerations [109, 773].

with $\gamma(R)$ the redshift factor. Thus, when we consider two clocks at fixed, but different radii R_1 and R_2 from the black hole, there will be a relative time dilation between them.

We introduce a particle field that interacts with the gravitational field of the black hole, and later we will measure the particle field with an Unruh–DeWitt detector. Specifically, we consider a massless scalar field $\phi(x)$ conformally coupled to the metric with equation of motion

$$\Box\phi - \frac{1}{8}R\phi = 0, \tag{9.3}$$

with vacuum analogous to the Hartle–Hawking state [109, 746].

The measurement probabilities of the Unruh–DeWitt detector, as explained below, depend on the two-point correlator of the scalar field in the BTZ black hole background. Such correlator can be calculated from the AdS_3 correlators (see [119, 653, 654, 746] for details):

$$W_{\mathrm{BTZ}}\left(\mathrm{x},\mathrm{x}'\right) = \sum_n \Upsilon^n W_{AdS}\left(\mathrm{x},\Gamma^n\mathrm{x}'\right) \tag{9.4}$$

with $\Upsilon = +1, -1$ for an untwisted and twisted field [746]. We will consider the untwisted case $\Upsilon = 1$ unless stated otherwise. The correlation function in the scalar vacuum of AdS_3 is given by

$$W_{\mathrm{AdS}}\left(x,x'\right) = \frac{1}{4\pi l\sqrt{2}}\left[\frac{1}{\sqrt{\rho\left(x,x'\right)}} - \frac{\zeta}{\sqrt{\rho\left(x,x'\right)+2}}\right], \tag{9.5}$$

with $\rho(x,x')$ the squared geodesic distance in the 2+2 embedding space of AdS_3, and $\zeta \in \{-1,0,1\}$ specifies the boundary conditions at asymptotic infinity: Neumann, transparent or Dirichlet [746, 763, 776]. We will take transparent boundary conditions $\zeta = 0$, unless stated otherwise. Neumann and Dirichlet boundary conditions are reflective boundary conditions, implemented through requirements on the derivative of the field and the field itself, respectively, when viewing the covering space of AdS as half of an Einstein static universe [746, 777]. In elementary wave-mechanical analogy of a free and a fixed endpoint, Neumann and Dirichlet boundary conditions lead to an in-phase and out-of-phase reflection of a wave. Transparent boundary conditions for a massless conformally coupled scalar field on (the covering space of) AdS, which is conformally isometric to half of an Einstein static universe, arise when taking the quantization of the Einstein static universe with no reflecting boundary conditions on AdS [777]. This leads to effective two-time Cauchy surfaces in the case of transparent boundary conditions for AdS and a recirculation of energy and angular momentum across conformal infinity [777]. As the (eternal) BTZ black hole is a quotient of an open region of AdS_3, the implications

of transparent boundary conditions for specifying initial data and conservation or flux-balance laws in the BTZ spacetime are not immediately clear and are beyond the purpose of this work; however, the BTZ two-point function for a massless conformally coupled scalar field can be calculated from the AdS_3 two-point function through image sums as explained above in Eq. (9.4). Other boundary conditions are possible as well; for instance, Refs. [754, 778] consider Robin-type boundary conditions for the BTZ spacetime. We refer to Appendix B.2 for more details on the BTZ two-point correlation functions.

9.2.2 Unruh–DeWitt detectors

In flat spacetime, the usual definition of a particle relies on excitations of fields, decomposed into positive and negative frequency modes — a decomposition that is possible when a preferred notion of time exists. On a general curved spacetime, nevertheless, this straightforward definition is not possible, due to the lack of a preferred notion of time. A possible way of defining particles in a more operational way was introduced by Unruh [779], suggesting that "a particle is what a particle detector detects".

Unruh modeled a particle detector as a scalar field in a cavity, and showed what we now call the Unruh effect [780]: an accelerated detector in Minkowski spacetime registers a thermal bath of particles. DeWitt [781] proposed a simplified detector model, replacing the scalar field in a cavity by a simple two-level system, obtaining the Unruh–DeWitt detector [197, 782].

Specifically, an Unruh–DeWitt (UDW) detector consists of a detector modeled as a quantum system following a path $x(\tau)$ in some curved spacetime that interacts linearly with a scalar field.[2] The spacetime path[3] can be freely chosen, geodesic or non-geodesic driven by some external force. The detector has internal degrees of freedom, typically taken to be an n-level system with energy gaps, that "click" when a particle is measured, and external degrees of freedom, corresponding to the position of the detector. It interacts with the particle field through a linear coupling, with the Schrödinger-picture interaction

[2] Extensions to detectors for fermionic or other boson fields exist [783, 784], as well as explicit models for n-level detectors [785, 786], detectors with quantized center of mass [670, 758], second-quantized detectors [756, 759] and detector models derived relativistic-covariantly using the influence functional [787]. For simplicity, we will use a 2-level detector model.

[3] Usually the spacetime path is some classical trajectory on a background spacetime; later, we will allow the detector in a superposition of paths (meaning the path is an operator with a state highly peaked around two classical trajectories). We will denote the corresponding coordinate position operator that controls the paths by the hatted version $\hat{\mathbf{x}}$.

Hamiltonian in the detector's rest frame being

$$\hat{H}^{\tau}_{\text{int}} = \lambda\eta(\tau)\hat{\mu}\otimes\phi(\mathbf{x}(\tau)), \tag{9.6}$$

with $x(\tau)$ the detector's prescribed spacetime trajectory and $\hat{\mu}$ the monopole moment which has off-diagonal elements in the detector's energy eigenbasis so that this interaction allows the detector to switch between its internal energy levels. Here $\hat{H}^{\tau}_{\text{int}}$ is written with respect to the proper time of the detector; choosing a different time t to describe the evolution yields the transformed Hamiltonian [755]

$$\hat{H}^{t}_{\text{int}} = \frac{d\tau}{dt}\hat{H}^{\tau}_{\text{int}}. \tag{9.7}$$

The coupling strength λ will be assumed to be small, and $\eta(\tau)$ is a switching (-on and -off) function of the detector, which we will take to be a Gaussian $\eta(\tau) = e^{-\frac{\tau^2}{2\sigma^2}}$ with variance much smaller than the timescale $\Delta\tau$ of the experiment. We say the detector has detected a particle when the detector transitions to a higher internal energy level. An example of a detector could be an atom as a detector for particles of the electromagnetic field.

In the remainder of this chapter we will also use an idealized quantum-controlled extension of this Unruh–DeWitt detector model [661–663] in a BTZ spacetime, in which the spatial coordinate specifying the prescribed detector path can be in quantum superposition. We denote the corresponding (spatial) coordinate-position operator controlling the paths by $\hat{\mathbf{x}}_D$, with spectral measure $E_{\hat{\mathbf{x}}_D}(d\mathbf{x}_D)$ so that $\hat{\mathbf{x}}_D = \int \mathbf{x}_D E_{\hat{\mathbf{x}}_D}(d\mathbf{x}_D)$, also denoted formally as $|\mathbf{x}_D\rangle\langle\mathbf{x}_D|\, d\nu(\mathbf{x}_D)$. We take the different trajectories to be prescribed externally and neglect the dynamics of the control degree of freedom and apparatus maintaining them; see for instance Refs. [670, 757–759] for treatments including dynamics of the quantized centre-of-mass. We denote by $x_{\mathbf{x}_D}(\tau)$ the prescribed spacetime trajectory associated with the spatial-coordinate value $\mathbf{x}_D$, which we will take to be static trajectories $x_{\mathbf{x}_D}(\tau) = (t_{\mathbf{x}_D}(\tau), \mathbf{x}_D)$ held fixed spatially at $\mathbf{x}_D$; see also Section 9.3.2 for a treatment modeling the reference fields explicitly. The quantum-controlled extension of the Unruh–DeWitt interaction becomes [661–663, 670, 757–759] (with pre-subscript I for interaction picture)

$${}_I\hat{H}^{\tau}_{\text{int}}(\tau) = \int E_{\hat{\mathbf{x}}_D}(d\mathbf{x}_D)\otimes\lambda\eta(\tau)\hat{\mu}(\tau)\otimes\hat{\phi}\left(x_{\mathbf{x}_D}(\tau)\right) \tag{9.8}$$

with $\hat{\mu}(\tau) = e^{i\hat{H}_0\tau}\,\hat{\mu}\,e^{-i\hat{H}_0\tau}$ with $\hat{H}_0$ the detector's monopole moment Hamiltonian so that $\hat{H}_0\left|E\right\rangle_{\mu} = E\left|E\right\rangle_{\mu}$, and τ the detector's proper time. We will denote states of the

detector's position and internal degrees of freedom by $|\mathbf{x}_D\rangle_D, |E\rangle_\mu$, respectively. For finitely many branches for $\mathbf{x}_D$ the integral reduces to a sum; for instance, for two sharply localized branches $|\mathbf{x}_1\rangle, |\mathbf{x}_2\rangle$ this reduces to

$$\sum_i |\mathbf{x}_i\rangle \langle \mathbf{x}_i|_D \otimes {}_I\hat{H}^\tau_{\text{int},i}(\tau), \tag{9.9}$$

the quantum-controlled form also used in Refs. [661, 788], with our control states having physical (coordinate) position labels, and ${}_I\hat{H}^\tau_{\text{int},i}(\tau) = \lambda\eta(\tau)\hat{\mu}\otimes\phi(x_i(\tau))$. Here $x_i(\tau)$ is the detector worldline of the detector held externally fixed at spatial $\mathbf{x}_i$, i.e., $x_i(\tau) := x_{\mathbf{x}_i}(\tau)$. In our BTZ set-up of this chapter, we will take all paths to be static with equal fixed angular coordinates, and the radial part of $\mathbf{x}_D$ will be the relational radial coordinate of the detector in a reference frame adapted to the black hole, or alternatively, under a QRF transformation, a radial coordinate indicating the black hole position in a frame adapted to the detector; Section 9.3.2 gives this control variable a physical interpretation using quantum reference fields.

This detector can be used to measure properties of the scalar field to gain information on the semiclassical gravitational field. To explain how a detector 'clicks', let us consider one single classical trajectory $x(\tau)$ of the detector, with the detector initialized in its ground state and field ϕ in the Hartle–Hawking vacuum state $|0\rangle_\phi$. The detector "clicks" when a transition to another internal state $|E\rangle_\mu$ happens, due to the interaction with the particle field. This leads to a transition amplitude (see Appendix B.3 for details)

$$\begin{aligned}\langle\vartheta, E|\psi(\tau_f)\rangle = i\lambda\langle E|\mu|E_0\rangle_\mu \times \\ \times \int_{\tau_i}^{\tau_f} d\tau\eta(\tau)e^{i(E-E_0)\tau}\langle\vartheta|\phi(x(\tau))|0\rangle_\phi,\end{aligned} \tag{9.10}$$

with $|\vartheta\rangle$ a final scalar field state that we will sum over, and later we will take $\tau_i = -\tau_f$. The probability of the transition $E_0 \to E$ happening, independently of the final state of the field, is

$$\begin{aligned}P &= \int d\vartheta\, |\langle\vartheta, E|\psi(\tau_f)\rangle|^2 \\ &= |\langle E|\mu|E_0\rangle|^2\lambda^2\mathcal{F}(E-E_0),\end{aligned} \tag{9.11}$$

with the detector response function given by

$$\begin{aligned}\mathcal{F}(\Omega) = &\int_{\tau_i}^{\tau_f} d\tau \int_{\tau_i}^{\tau_f} d\tau' \eta(\tau)\eta(\tau') e^{-i\Omega(\tau-\tau')} \\ &\langle 0|\phi(x(\tau))\phi(x(\tau'))|0\rangle \\ &= \int_{\tau_i}^{\tau_f} d\tau \int_{\tau_i}^{\tau_f} d\tau' \eta(\tau)\eta(\tau') e^{-i\Omega(\tau-\tau')} \\ &W_{BTZ}(x(\tau), x(\tau')).\end{aligned} \tag{9.12}$$

Here $W_{BTZ}(x, x')$ is the scalar field two-point correlator in a BTZ black hole background.

In Sections 9.3 and 9.4 we will specialize this quantum-controlled detector construction to branches in which the detector is held at a fixed spatial coordinate $\mathbf{x}_i$, so that each branch corresponds to a prescribed static non-geodesic trajectory.

9.2.3 Quantum Reference Frames

When describing physical systems, we typically do so relative to a chosen reference frame. A compelling question emerges when considering the possibility that the reference itself exhibits quantum properties. This idea has sparked significant research into quantum reference frames [328, 329, 464, 471, 472, 683, 732–736], which investigate how the physical world is described when taking the reference frame to be a quantum system. Notably, it has been shown that fundamental quantum features, such as superposition and entanglement, depend on the chosen frame [329, 734]: a system that appears to be in superposition with respect to one quantum reference frame may be localized in another. This idea has been especially valuable in the study of quantum-gravitational phenomena, as it allows for a description within a reference frame where gravity is entirely classical. As a result, it offers insights into such phenomena without requiring a complete theory of quantum gravity.

In order to introduce these ideas, let us first consider three classical physical systems, B, D, and F as particles on a Cartesian spatial line, with the usual translation structure available. From the reference frame of B, we say that D and F are at position $\mathbf{x}_D$ and $\mathbf{x}_F$ respectively. On the other hand, when taking D as a reference, one assigns positions $\mathbf{q}_B = -\mathbf{x}_D$ and $\mathbf{q}_F = \mathbf{x}_F - \mathbf{x}_D$ to B and F respectively. Can we extend this notion to quantum systems? Is it possible to define a transformation between *quantum* reference frames?

If the three systems are now quantum, it could be that D is in a superposition state of two locations $\mathbf{x}_1$ and $\mathbf{x}_2$. How can we transform to the reference frame of D in this

scenario? A natural idea is to extend the classical transformations by linearity, namely applying the classical transformation for each branch of the superposition. The previous situation, from the reference frame of B, is described by

$$|\psi\rangle^{(B)} = \frac{1}{\sqrt{2}}\Big(|\mathbf{x}_1\rangle_D + |\mathbf{x}_2\rangle_D \Big) |\mathbf{x}_F\rangle_F \,. \tag{9.13}$$

A QRF transformation, i.e. a coherent translation controlled on system D's position, is implemented through the unitary transformation

$$\hat{S}^{(B)\to(D)} = \int d\mathbf{x} \; |-\mathbf{x}\rangle_B \langle\mathbf{x}|_D \otimes T_F^\dagger(\mathbf{x}), \tag{9.14}$$

with T_F the translation operator on the Hilbert space of the system F. The state in the perspective of D is then

$$\begin{aligned} |\psi\rangle^{(D)} &= S^{(B)\to(D)} |\psi\rangle^{(B)} \\ &= \frac{|-\mathbf{x}_1\rangle_B |\mathbf{x}_F - \mathbf{x}_1\rangle_F + |-\mathbf{x}_2\rangle_B |\mathbf{x}_F - \mathbf{x}_2\rangle_F}{\sqrt{2}}. \end{aligned} \tag{9.15}$$

This describes the same physical setup but when choosing the origin to be system D's location[4], see also Fig. 9.2.

Although the introductory example above uses ordinary spatial translations, the analogous QRF map $\hat{S}^{(B)\to(D)}$ used later in this work in the BTZ spacetime is a branch-dependent change of coordinate reference-field coordinates, making use of diffeomorphism invariance rather than a translation symmetry (there is no radial translation symmetry) or isometry of the BTZ geometry.

Consider now the situation where not only we want to (spatially) 'translate' the physical description, with a translation defined in the coordinate space using reference fields of Section 9.3.2, but also change the time coordinate we use in Schrödinger's equation. To do so, the (classical) reference frame change involves two steps: we first apply the translation, and then transform the Hamiltonian to account for the new time parameter. Under this change of classical reference frame, the Hamiltonian transforms as

$$\hat{H}^{(D)} = \frac{dt}{d\tau} T \hat{H}^{(B)} T^\dagger, \tag{9.16}$$

where T implements the classical spatial translation, τ is the new time parameter and t is the old one. Applied to our set-up, if we choose t to be the coordinate time, and τ to

[4]Sometimes the origin state is also written explicitly, i.e. for example $|\mathbf{0}\rangle_D$ in this case.

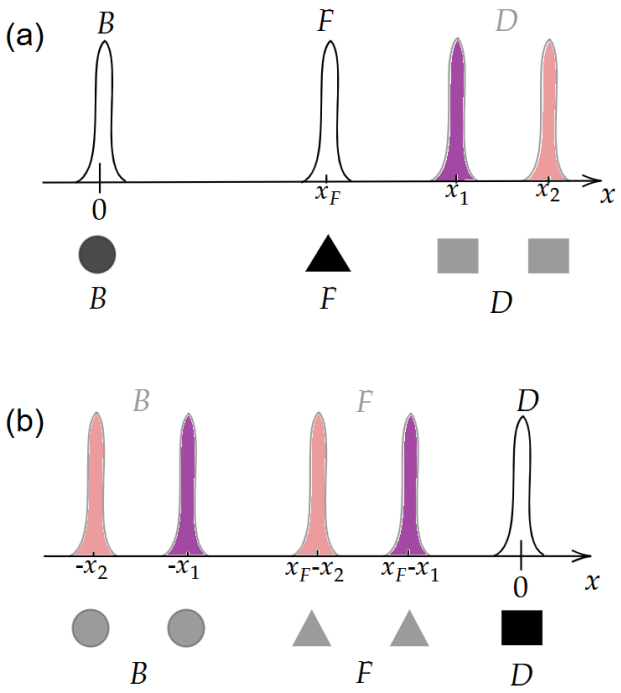

Figure 9.2: (a): Systems B and F, near a position-superposed system D, which under a QRF transformation can be mapped to (b), where B and F are in indefinite, entangled locations but D has a definite single-valued location. Adapted from Ref. [329].

be the proper time in the detector frame, we obtain

$$\tau = \gamma(\mathbf{x}_D)\, t, \tag{9.17}$$

$$\gamma(\mathbf{x}_D) = \sqrt{-g_{00}(\mathbf{x}_D)} = \sqrt{\frac{\mathbf{x}_D^2}{l^2} - M}. \tag{9.18}$$

We now generalize this transformation to superposition of translations, as we have explained before. The resulting Hamiltonian is (see for instance Ref. [756] for a similar treatment)

$$\hat{H}^{(D)} = \hat{S}^{(B)\to(D)} \gamma^{-1}(\hat{\mathbf{x}}_D) \hat{H}^{(B)} \hat{S}^{\dagger(B)\to(D)} \tag{9.19}$$

The reparametrization of time is now promoted to an operator: when D is in superposition with respect to B, their proper times are in superposition with respect to each other. In the rest of this chapter we apply these ideas to the quantum-controlled extension of the Unruh-DeWitt detector scenario described in the previous section, involving a detector D, a black hole B and a scalar field ϕ. The operator $\hat{\mathbf{x}}_D$ is the same coordinate-position/control operator controlling the paths as introduced in the previous section, and the interaction depends on the field through the local operator $\phi(\hat{x}_D)$, namely the field

evaluated at the detector position. Since ϕ is a scalar, the change of frame introduces no additional internal or tensorial transformation on the field degrees of freedom. Its only effect is the corresponding change of the argument at which the field is evaluated, which is already accounted for by the transformation of the detector position. Therefore, the QRF transformation acts non-trivially only on the relative position degrees of freedom of B and D, while acting as the identity on the field Hilbert space:

$$\hat{S}^{(B)\to(D)} = \int d\nu(\mathbf{x}) \; |-\mathbf{x}\rangle_B \langle \mathbf{x}|_D \otimes \mathbb{1}_\phi. \tag{9.20}$$

We will provide a realization of this transformation using reference fields in Section 9.3.2, which makes the relational meaning of the coordinate-position operator and the branchwise coordinate map explicit.

9.3 Set-up, QRF transformation and reference fields

9.3.1 Setup

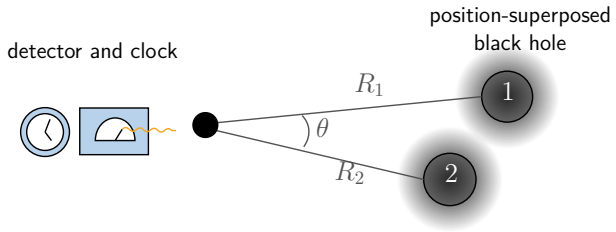

Figure 9.3: A detector with a clock is used to measure scalar particles created some fixed distance away from a position-superposed black hole. Relative to the detector, the black hole is in a superposition of radial distances R_1 and R_2.

Setup. Consider the following scenario, illustrated in Fig. 9.3. A particle detector is used to measure a massless conformally coupled scalar field near a 2+1d BTZ black hole that can be interpreted as being in a superposition of locations as viewed from a reference frame adapted to the detector. Such a black hole can, for example, result from the collapse of a shell in a superposition of two different positions in this frame. The detector comes with a clock and is held stationary rather than freely falling; this is the particular prescribed non-geodesic trajectory considered here, maintained by an external force whose dynamics we neglect. The detector is turned on for a finite proper time interval according to its clock.

How does the detector respond to this situation? Can it witness quantum effects of the superposed spacetime?

Approximations. In our calculations, we neglect the gravitational backreaction of the scalar field and detector, as well as reaction effects of the detector's clicking and we assume the two 'translated' frame-dressed black hole metrics to be orthogonal[5]. We also assume the black hole metric to be static, i.e. we neglect potential evaporation. As introduced in Section 9.2.2, we work with a point-like detector, i.e. we assume the detector to be highly localized in each branch of the evolution. Moreover, we also neglect a potential conjugate variable to the black hole position [452, 453, 609].

Using QRFs. In order to model the setup described above, we cannot employ the usual treatment of BTZ black holes and Unruh–DeWitt detectors: since the black hole is in superposition of locations in the chosen detector's reference frame, the evolution of the particle field will be in a quantum superposition of spacetimes, and the usual description of QFT on curved spacetime does not suffice to describe this situation.

Despite not subscribing to any specific theory of quantum gravity, we assume here that any such theory must have a semiclassical limit, where perfectly distinguishable gravitational states are orthogonal, and such that in each branch classical general relativity applies. We assign to each of these states a vector on a Hilbert space $\mathcal{H}_B = L^2(\mathbb{R})$, labeled by the position of the black hole with respect to the detector.

Under these assumptions, we can model the physical setup in the perspective of the detector, before it is switched on, as

$$|\psi(\tau_i)\rangle^{(D)} = \frac{|\mathbf{x}_1\rangle_B + |\mathbf{x}_2\rangle_B}{\sqrt{2}} |0\rangle_\phi |E_0\rangle_\mu \,. \tag{9.21}$$

Here $\left|\mathbf{x}_{1/2}\right\rangle_B$ are the two orthogonal semiclassical states of the black hole, corresponding to two different coordinate positions relative to the detector (see also Section 9.3.2) with $|0\rangle_\phi$ the vacuum state of the scalar field in the gravitational BTZ background, evolved according to standard QFT on curved spacetime, and $|E_0\rangle_\mu$ represents the internal degrees of freedom of the detector before it is switched on.

In order to model the response of the detector, we need to define the Hamiltonian under which it evolves in this situation. A possible solution is to postulate it, as was

[5] Or more precisely, we assume different eigenstates for the operator specifying the relational radial coordinate distance between the detector and black hole to be orthogonal, see also Section 9.3.2.

done in Ref. [119]. We follow here an alternative approach, using QRF transformations to derive it.

As described in the previous section, the physical setup of the black hole in superposition of locations with respect to a detector can be described in the reference frame of the black hole as the detector being in superposition of locations and the black hole being localized (cf. Fig. 9.2). In other words, by applying the QRF change defined in Eq. (9.20), we obtain the state

$$|\psi(t_i)\rangle^{(B)} = \frac{|-\mathbf{x}_1\rangle_D + |-\mathbf{x}_2\rangle_D}{\sqrt{2}} |0\rangle_\phi |E_0\rangle_\mu . \tag{9.22}$$

where we choose the initial coordinate time to be $t_i = \tau_i = 0$. In this reference frame, there is no source of gravitational field in quantum superposition, and hence we can employ the usual tools of QFT in curved space time. Specifically, the UDW interaction for a localized black hole and quantum-controlled detector position introduced in Section 9.2.2 has the following form with respect to the coordinate time [756]:

$$H_{\text{int}}^{(B),t} = \gamma(\hat{\mathbf{x}}_D)\lambda\, \eta(\gamma(\hat{\mathbf{x}}_D)t)\, \mu \otimes \phi(\hat{\mathbf{x}}_D). \tag{9.23}$$

Hence we can perform the QRF transformation Eq. (9.20) to obtain the Hamiltonian in the detector's frame with respect to its proper time as

$$H_{\text{int}}^{(D),\tau} = \lambda\, \eta(\tau)\mu \otimes \phi(-\hat{\mathbf{x}}_B), \tag{9.24}$$

see also Section B.5. This Hamiltonian is analogous to the one that was postulated in Ref. [119], in the case of a mass-superposed black hole.

9.3.2 A treatment of the position-superposition using reference fields

This section presents a reference-field realization of the restricted quantum reference frame construction used above. It provides a relational and diffeomorphism-invariant interpretation of the coordinate-position operator $\hat{\mathbf{x}}$ appearing in the quantum-controlled UDW interaction and derives the branchwise coordinate relabelling underlying Eq. (9.20). This section is not intended as a complete dynamical treatment of quantum reference fields, and operates within the approximations stated below. Readers not interested in these details can skip this section. In subsequent later sections we comment on when and how this introduction of reference fields changes certain calculations or expressions.

In the previous section we described the set-up of a detector-black hole system where the relative coordinate distance between the detector and black hole is in a

superposition. To give an operational meaning to such a position superposition, and to clarify the meaning of the radial specialization of the control operator $\hat{\mathbf{x}}$ as a relational radial coordinate distance between the black hole and detector, we introduce coordinate reference fields [471, 789–791]. This can be seen as a more precise kinematical treatment of quantum reference frames as outlined in the previous section, and clarifies the meaning of the coordinate operator $\hat{\mathbf{x}}$ in the Unruh–DeWitt interaction and its QRF transform. While we restrict to our specific set-up of two quantum reference fields restricted to specific sets of states, a quantum-controlled detector worldline, a fixed metric and conformally coupled scalar field and no backreaction, seeking a QRF transformation that works in this limited setting, a more complete quantum treatment of QRFs and their transformations is an active topic of research in literature; we comment on what a more complete dynamical treatment may involve, and refer the reader to Refs. [789, 791, 792] for work in this direction. We comment upon this in the paragraph *More complete treatment* below.

QRFs of reference fields. In 2+1 dimensions, a classical frame of reference fields can be given by a locally invertible map $X^A : U \subset M \to X(U) \subset \mathbb{R}^3$ of three reference fields X^0, X^1, X^2 that we take to be three scalar fields. In a complete quantum treatment of reference fields, a quantum coordinate reference frame consists of quantum scalar fields, restricted to a subspace formally spanned by field eigenstates with eigenfunctions that define locally valid coordinates. In our limited treatment, we will restrict the fields to states $|\Psi_X\rangle$ sharply peaked around a classical field configuration X^A with negligible fluctuation $\delta\hat{X}^A = \hat{X}^A - X^A$. We will nevertheless allow for these quantum reference fields to be in states sharply peaked around a superposition of classical field configurations, i.e. $|\Psi_{X,1}\rangle + |\Psi_{X,2}\rangle$. We will neglect the backreaction of the reference fields on all other systems including the black hole spacetime, and will assume factorization of the different systems and fields involved.[6]

A reference frame can be used to dress fields and observables into gauge-invariant operators (i.e. invariant under diffeomorphisms on M) [526, 789, 793–795]. For instance, given a spacetime scalar F, its dressing by X has value $F_X(\xi^A)$ at the physical coordinate reading $X^A = \xi^A$ given by $F_X(\xi^A) = F(X^{-1}(\xi^A))$. Note that F_X is indeed invariant under diffeomorphisms on M, as both F and X transform. In the case of a quantum reference frame $\hat{X}^A$, such dressing can be defined formally

[6] For a treatment in linearized gravity that does not neglect backreaction see Ref. [789].

through its eigenfunctions[7] $|X\rangle_X$, $F_{\hat{X}}(\xi)\,|X\rangle_X = F(\hat{X}^{-1}(\xi))\,|X\rangle_X = F(X^{-1}(\xi))\,|X\rangle$. The inverse $\hat{X}^{-1}$ is thus defined branchwise, applying the inverse of the classical field configuration, X^{-1}, in each spectral branch. A similar construction holds for spacetime tensors, with the X-dressing obtained by pulling back the tensor along X^{-1}; for instance, $(g_X)_\xi(u,v) = (X^{-1})^* g_(\xi)(u,v)$. The metric g on M that we will consider is classical (but its dressing might change (quantumly) under (quantum) dressings).

The detector worldline. The detector's path $z : I \subset \mathbb{R} \to M : \tau \to z(\tau)$[8] will be presented in a specific reference frame by specifying $z^{(X)} := X \circ z : I \subset \mathbb{R} \to \mathbb{R}^3$. We will allow $z^{(X)}$ to be a quantum operator $\hat{z}^{(X)}$ with, formally, eigenfunctions denoting detector worldlines $z^{(X)}$. In particular, we will take the radial coordinate position of the detector z_r to be an operator, and consider detector paths that remain at fixed values of the spatial parts X^1, X^2 of the reference frame. This radial component $\hat{z}_r$ is the radial specialization of the coordinate-position/control operator $\hat{\mathbf{x}}_D$ introduced in Section 9.2.2. We take the detector path to be externally prescribed, but potentially in a quantum superposed semiclassical state, neglecting quantum fluctuations around the semiclassical states, while ignoring all dynamics related to this external force or holding apparatus and path-control degree of freedom. Therefore, it can be seen as a detector model with quantum-controlled prescribed trajectories of Refs. [661–663]; see Refs. [757–759] for treatments taking into account the dynamics of a detector with quantized center of mass.

We will consider reference frames X_B, adapted to the black hole, and $Y_{D,i}$, adapted to the detector in branch i, and construct a quantum reference frame transformation between them, which can be seen as the equivalent of Eq. (9.20).

QRF adapted to the black hole. The first reference frame X_B^0, X_B^1, X_B^2 that we consider is one adapted to the BTZ black hole exterior, and in the coordinate system t, r, φ of Eq. (9.1), we simply take these to be

$$X_B^0(x) = t(x), \quad X_B^1(x) = r(x), \quad X_B^2(x) = \varphi(x). \tag{9.25}$$

This can be seen as an eigenfunction in the formal eigenstate $|X\rangle$ of the operator $\hat{X}$. In this treatment of reference fields, we do not require our reference fields to satisfy

[7] One can define such eigenfunctions and eigenvectors using equal-time hypersurfaces in an auxiliary foliation [789]. In the case of quantum gravity with superposed metrics, for instance, such a foliation and timelike-spacelike split could be defined branch per branch.

[8] Here $z(\tau)$ denotes the underlying spacetime worldline; its coordinate representation is the $x(\tau)$ used in Section 9.2.2.

a prescribed equation of motion (and thus they can be considered 'idealized' or 'test' reference fields). If, instead, for instance, we wanted our reference fields to be massless scalar fields on the BTZ spacetime satisfying $\Box X^A_B = 0$, we would have to change X^1_B, for example taking $X^1_B(x) = C_1 \ln|1 - \frac{r_h^2}{r(x)^2}| + C_2$, for some constants C_1, C_2.

Our choice of reference frame X means the metric in X-coordinates (the X-dressed metric g_X) is given by

$$\mathrm{d}s^2 = -f(r)\,\mathrm{d}t^2 + f(r)^{-1}\,\mathrm{d}r^2 + r^2\,\mathrm{d}\varphi^2, \tag{9.26}$$

with $f(r) = r^2/l^2 - M$.

Specifying the detector in the X-frame. To specify the detector location, we start in the X-reference frame, and establish the detector location through the reference frame X, considering $z^{(X)}(\tau) = X \circ z(\tau)$. All detector trajectories we consider take the detector to be held externally fixed at a certain radial and angular coordinate, rather than freely falling. We will nevertheless allow our detector location to be in a superposition of radial locations $X^1 = r_{D,1}$ and $X^1 = r_{D,2}$. Therefore, the radial component of $z^{(X)}$ is promoted to an operator denoted $\hat{z}_r(\tau)$. We will write its value in branch i as $r_{D,i}$. Formally, $\hat{z}_r$ is an operator acting on some control system that controls the path of the detector (keeps it fixed at some quantum radial location), and so acts on the Hilbert space of a particle on the coordinate line r, i.e. $\mathcal{H} = L_2(\mathbb{R}, dr)$ (or more precisely, taking an open part $I \subset \mathbb{R}$, $L_2(I, dr)$). In this reduced representation its spectral measure will be denoted $E_{\hat{z}_r}(dr)$, also written formally as $E_{\hat{z}_r}(dr) = |r\rangle\langle r|dr$.

The relational radial coordinate position between the detector (held fixed at certain radial and angular coordinate) and black hole can, for instance, be given meaning through the redshift factor on the detector worldline $z_i(\tau)$,

$$\frac{d\tau}{dX^0(z_i(\tau))} = \gamma(R_i), \tag{9.27}$$

which is a function of the radial coordinate on that worldline. Other ways to give it an operational meaning include using signals sent from a clock at the boundary.

Witnessing branch coherence. To certify that the detector is in a superposition of worldlines, or under a QRF transformation, the black hole in a superposition of radial locations as seen from a frame adapted to the detector, one possible way to witness this coherence is to recombine the branches and perform a measurement that erases which-branch information, similarly to the $|\pm\rangle$ measurement of Section 9.4.1, or, more generally, perform a measurement sensitive to the off-diagonal coherence. In fact, the

interference of Section 9.4.1 can be viewed as constituting such superposition witness as it is absent in the case of the corresponding incoherent classical mixture. Operationally, the measurement and initial preparation themselves may require additional control systems. If these require additional quantum reference systems but the corresponding operations employ idealized absolute quantities, the localizability of the relevant reference quantities (represented in general by POVMs) and the localization of the reference system's state determine how accurately the idealized measurement and state preparation can be implemented [735, 793, 796]. Additionally, repeated use of the same finite-resource reference system may degrade its state and consequently its suitability as a (good) reference [797, 798]. We will not go deeper into these issues here.

QRF adapted to the detector. The second reference frame Y^0_D, Y^1_D, Y^2_D that we consider is a reference frame adapted to the detector, meaning it is well-localized with respect to the detector, taking in branch i

$$Y^0_{D,i}(x) = t(x), \quad Y^1_{D,i}(x) = r(x) - r_{D,i}, \quad Y^2_{D,i}(x) = \varphi(x), \tag{9.28}$$

with t, r, ϕ the BTZ coordinates of Eq. (9.1), with detector at $Y^1_{D,i}(x) = 0, \phi(x) = 0$. Note that the condition $Y^1_{D,i}(x) = 0$ means that $r(x) = r_{D,i}$ with $r_{D,i}$ the (static) radial location of the detector in the t, r, ϕ-coordinates in branch i (recall that the detector is held fixed at some fixed radial distance in each branch).

Note that we could as well have adopted the choice to instead consider the first reference field $Y^0_{D,i}$ (timelike in the black hole exterior that we are probing) to be a reference field that reduces to the detector's proper time at the detector's location (and is appropriately extended away from it to the whole open region considered), i.e. $Y^0_{D,i}(x) = \tau(x)$ with $\tau(x)|_{\text{detector location}} = \tau$. Instead, we have opted to implement the additional redshift factor from relating the time coordinate and proper time of the Hamiltonians in each frame, as presented in Section 9.2.3. Had we chosen this detector's proper time coordinate extension as the reference field $Y^0_{D,i}$, this branch-dependent redshift factor would result from the QRF transformation between these two frames of reference fields X_B and $Y_{D,i}$ in each branch outlined below.

An alternative to two different reference fields in the two branches is to consider an operator $\hat{Y}$ restricted to a subspace generated by eigenvectors $|Y\rangle$ with eigenfunctions of the form $Y_{D,i}$ for some $r_{D,i}$, with $r_{D,i}$ the eigenvalues of $\hat{z}_r$, the radial component of the detector worldline in the X-frame $X \circ \hat{z}$.

Note that the two branches are not gauge-related because of the different relational distance $r_{D,i} = X_B^1(z_i)$ between the two branches.

Transforming between QRFs of reference fields. The classical reference frame transformation between these two frames in branch i is determined by $X_B \circ Y_{D,i}^{-1}$, loosely stated, determined by a change of reference coordinate fields given by $r \to r + r_{D,i}$ in each branch. Now let us make this more precise.

Classically, a transformation from reference frame fields X^A to Y^A is determined by

$$Q_{X\to Y} : \mathbb{R}^3 \to \mathbb{R}^3 : \xi^A \to X \circ Y^{-1}(\xi), \tag{9.29}$$

acting on dressed configuration space operators by its pullback. For instance, the spacetime scalar F with $F_X : \mathbb{R}^3 \to \mathbb{R}$ defined by $F_X(\xi) = F(X^{-1}(\xi))$ does transform correctly under this pullback:

$$\begin{aligned} F_Y(\xi) &= F(Y^{-1}(\xi)) = F(X^{-1} \circ X \circ Y^{-1}(\xi)) \\ &= F\Big(X^{-1}(X \circ Y^{-1})(\xi)\Big) = F_X(X \circ Y^{-1}(\xi)) \\ &= Q^*_{X\to Y} F_X(\xi) \end{aligned} \tag{9.30}$$

If, in our superposed BTZ case, we now apply this transformation coherently in each branch i, we get in branch i the map

$$Q_{X_B\to Y_D,i} = X_B \circ Y_{D,i}^{-1} : \xi^A \to \xi^A + \delta^A_r\, r_{D,i}. \tag{9.31}$$

We can thus make this work in each branch by considering the operator

$$\hat{Q}_{X\to\hat{Y}} : \xi^A \to \xi^A + \delta^A_r\, \hat{r}_D^{(X)}, \tag{9.32}$$

with $\hat{r}_D^{(X)}$ the operator for the radial coordinate position between the detector and black hole.

In branch i in the reference frame adapted to the detector, the $Y_{D,i}$-dressed metric $g_{Y_{D,i}}$ can be written as

$$\mathrm{d}s^2 = -f_i(\overline{r})\,\mathrm{d}t^2 + f_i(\overline{r})^{-1}\,\mathrm{d}\overline{r}^2 + (\overline{r} + r_{D,i})^2\,\mathrm{d}\varphi^2, \tag{9.33}$$

with $f_i(\overline{r}) = (\overline{r} + r_{D,i})^2/l^2 - M$, with $t, \overline{r}, \varphi$ now denoting $Y^0_{D,i}, Y^1_{D,i}, Y^2_{D,i}$. Our set-up can thus be interpreted as a black hole in a position-superposition as seen from the Y-frame. The equations of motion are covariant under this change of frame.

Concluding, we thus see that this map $\hat{Q}_{X\to Y}$ and its pullback reproduce the branch-wise coordinate relabelling underlying the QRF transformation of Eq. (9.20):

$$\hat{S}^{(B)\to(D)} = \int d\nu(\mathbf{x})\ |-\mathbf{x}\rangle_B \langle \mathbf{x}|_D \otimes \mathbb{1}_\phi, \tag{9.34}$$

where the measure $d\nu(\mathbf{x})$ is a chosen measure for the reduced relational coordinate Hilbert space and could also be written equivalently using the spectral-measure notation of Section 9.2.2; it is not a spacetime volume measure. The identity on ϕ arises as the QRF transformation acts as the identity on the abstract field-state factor (the undressed field ϕ). Had we considered the dressed field ϕ_X, it would change accordingly, as explained above.

The QRF map $\hat{Q}_{X\to\hat{Y}}$ we found is a reduction of the full quantum reference frame transformation, only intended to work correctly in the limited set-up we consider. It arises from the operator nature of $\hat{Y}$, and more specifically, $\hat{Y}^1$ through the definition of the detector worldline $\hat{z}^{X^1}$ in the X-reference frame. Note that this treatment clarifies the meaning of the coordinate operator $\hat{\mathbf{x}}$ used in Eqs. (9.23) and (9.24) and the choice of assigning a Hilbert space $\mathcal{H}_B = L^2(\mathbb{R}, dr)$ to the black hole-detector system that captures this relational degree of freedom in the radial coordinate; it is the relational parameter describing the transformation between the reference fields, which happens to be a translation in the radial r-coordinate, $r \to r + r_{D,i}$, which is why in the main text we refer to this as a 'translation'. (It is thus not intended to refer to a translation symmetry as there is none in the BTZ spacetime.)

A more complete quantum reference frame transformation, not restricted to a specific set of reference field states, considers (pullbacks of) $\hat{Q}_{\hat{X}\to\hat{Y}} = \hat{X} \circ \hat{Y}^{-1}$, which can be defined (formally) through its eigenfunctions. Moreover, such QRF transformation may be defined through a perspective-neutral approach that takes into account the full dynamics and constraints [464, 733, 789, 792], see also the paragraph *More complete QRF treatment* below. The frames we consider are idealized, sharply localized frames within the restricted configuration subspace considered here, resembling the ideal-frame limit of operational QRF constructions [328, 733, 735, 736, 793], although we do not construct a covariant PVM or invariant measure for the full diffeomorphism group.

Connecting to literature. We can connect this treatment also to that of Ref. [471]. Namely, being in the reference frame of X_B, we write the initial state as

$$|X_B\rangle \left(|x_1\rangle_D \left|Y^A_{D,1}\right\rangle + |x_2\rangle_D \left|Y^A_{D,2}\right\rangle\right)/\sqrt{2}.$$

Going to the reference frame of Y_D, we perform a branch-dependent diffeomorphism $d^{(i)}$ on the manifold M (or the open region we care about), given by $r \to r + r_{D,i}$ in branch i, or more precisely,

$$\begin{aligned} d^{(i)} : M \to M : p = X_B^{-1}(t_p, r_p, \varphi_p) \\ \mapsto d^{(i)}(p) = X_B^{-1}(t_p, r_p + r_{D,i}, \varphi_p). \end{aligned} \tag{9.35}$$

This sets indeed $\left(d^{(1)}\right)^* (Y_{D,1}) = \left(d^{(2)}\right)^* (Y_{D,2}) = \tilde{Y}_{D*}$, so that we end up in a state where $\hat{Y}_D$ is single-valued (note that there are more diffeomorphisms that achieve this). The branch-dependent diffeomorphisms are related to the above QRF transformation through $X_B \circ d^{(i)} \circ X_B^{-1} = X_B \circ Y_{D,i}^{-1}$.

More complete QRF treatment. We have taken the reference fields to have zero backreaction, non-interacting with other fields and ignored their dynamics. In a more complete treatment, we should specify their dynamics and include their backreaction. Different formalisms then provide different counterparts of the reduced frame transformation we considered here.

In particular, the frame change map $\hat{X} \circ \hat{Y}^{-1}$ and the change of operator-dressing for the configuration variables may not determine the complete transformation; it may be necessary to further obtain its action on conjugate momenta, constraints, physical Hamiltonians, or the physical inner product. For instance, the perspective-neutral approach [146, 151, 464, 733] considers the physical space encoding all different QRF perspectives, from which one reduces to a QRF X's perspective through a reduction map R_X, with the QRF transformation given by $R_Y R_X^\dagger$, a unitary if the reduction maps are isometric isomorphisms onto the relevant reduced physical spaces. This idea can be implemented canonically [146, 151, 464, 570, 789, 799], through the path integral [792] and quantization of classical covariant phase-space methods [526, 800]. Other, related approaches are the operational or algebraic approach [683, 735, 736, 801] and the extra-particle approach [802], whose relation to the perspective-neutral approach remains an active topic of investigation [467, 803].

Moreover, we have neglected the action of the reference frame transformation on boundary data; in the presence of (asymptotic) boundaries, edge modes might appear if a boundary is held fixed [470, 743, 800].

9.4 Main results: detector response near a position-superposed black hole

9.4.1 Results

Measuring the final state. We now have a full description of the physical situation: given the initial state and interaction Hamiltonian in Eqs. (9.21) and (9.24), we can calculate the final state of the system for a transition to energy level E and final scalar field state ϑ (which we will sum over), to first order in the interaction coupling. The resulting state is simply a linear superposition, i.e.

$$\langle E, \vartheta | \psi(\tau_f)^{(D)} \rangle = \frac{|\mathbf{x}_1\rangle_B \, \Theta_1 + |\mathbf{x}_2\rangle_B \, \Theta_2}{\sqrt{2}}, \tag{9.36}$$

where $\Theta_{1/2}$ are the single-branch transition amplitudes defined in Eq. (9.10) for detector position at the origin and black hole at relational position $\mathbf{x}_{1/2}$.

In our limited treatment that neglects the gravitational dynamics of the black hole and the dynamics of the control variable or apparatus keeping the detector in place, the relative phase between the branches $\Delta\Phi$ is zero. In a treatment not neglecting these, an additional phase might arise. The relative phase $\Delta\Phi$ in our case may have several potential origins or contributions in a more complete treatment. First, it may arise from the approximate evolution of the black hole as an approximate eigenstate of the gravitational Hamiltonian with its mass as eigenvalue with respect to coordinate time t if we were to include the corresponding Hilbert space and state in our initial-state description. A relative phase difference when considering proper-time evolution might then appear as coordinate time relates differently to the detector's proper time in the two branches. The precise assumption in that case would thus be that for the purpose of the experiment the BTZ black hole spacetime remains approximately static and is an approximate eigenstate of the gravitational Hamiltonian, so that the phase coming from the black hole evolution is determined by its eigenvalue, presumably proportional to the black hole's ADM mass M (in line with [119]), i.e. $H_0^t \, |M\rangle_{BH} = M \, |M\rangle_{BH}$. Such phase might thus also be particularly relevant in the case of a mass-superposed black hole where a phase difference between the two branches of the superposition can also develop as the mass differs between the two branches. As we consider our black hole to have the same mass in both branches of the superposition, we do not assign an explicit Hilbert space for the mass to the black hole, only for its position. In a more

complete treatment, the spacetime location and energy-momentum of the black hole may not be independent [452, 609]. Approximating the black hole as developing only a phase independent of spatial location might also be an issue because of noncovariance, which would require a dynamical analysis taking into account the gravitational constraints and backreaction, potential non-factorization, etc. We neglect these complications here. Secondly, we also neglect any additional relative phase associated with the external control degree of freedom and the apparatus responsible for keeping the detector on the prescribed static trajectory. If one were to take into account these phases, we could also simply absorb them in the definition of the states. Whether an additional phase would be visible in the experimental data depends on the precise implementation of the interferometric black hole position measurement which we describe below.

In order to witness interference effects, we must lose which-way information. For this reason, a simple measurement on the internal transition of the detector from E_0 to E does not suffice, as it would produce a mixed signal $P_1 + P_2$ with no interference effect. Instead, we measure a transition to an energy eigenstate $|E\rangle$, together with a measurement on the BH position that ignores which-way (or which-branch) information[9] $|\pm\rangle_B = \frac{|\mathbf{x}_1\rangle_B \pm |\mathbf{x}_2\rangle_B}{\sqrt{2}}$. This is the standard measurement of interferometry, where the two branches are recombined with a beam splitter, and subsequently the output ports of the interferometer are measured. This may be hard to realize for a cosmological black hole, but might be possible experimentally in the context of analogue gravity [639, 804–807]. The measurement probabilities are

$$P_\pm = \frac{1}{4}\big(P_1 + P_2 \pm 2P_{12}\big), \tag{9.37}$$

where $P_{1/2}$ are the probabilities defined in Eq. (9.11) for a detector in a single black hole spacetime (but with the black hole at a different location $\mathbf{x}_1, \mathbf{x}_2$ relative to the detector). P_{12} is an interference term obtained as $P_{12} = \mathrm{Re}[F_1^* F_2]$, with F_i the single-branch amplitudes of Eq. (9.10), now summed over the final field states, for the black hole at

[9] If we model the physical reference fields of Section 9.3.2 as well, we would have to do a joint measurement that ignores which-branch information by including all systems S that (might) contain such which-branch information S_i, including the physical reference fields, and, in fact, also the device holding the detector at fixed radial location (which is not a geodesic), meaning $|\pm\rangle = 1/\sqrt{2}(|\mathbf{x}_1\rangle_B |S_1\rangle_S \pm |\mathbf{x}_2\rangle_B |S_2\rangle_S)$. If we did not do so, systems holding which-branch information might decohere the superposition leading to suppression of the off-diagonal term P_{12}.

relational positions $\mathbf{x}_1, \mathbf{x}_2$, that is

$$P_{12} = |\langle E|\mu|E_0\rangle|^2\lambda^2 \int_{\tau_i}^{\tau_f} d\tau d\tau' \eta(\tau)\eta(\tau') e^{-i\Delta E(\tau-\tau')} W^{12}_{BTZ}(x(\tau), x(\tau')) \tag{9.38}$$

with $\Delta E = E - E_0$ and $W^{(12)}_{BTZ}$ the two-point function

$$W^{(12)}_{BTZ}(x, x') = \langle 0_1|_\phi \phi_1(x)\phi_2(x')|0_2\rangle_\phi, \tag{9.39}$$

and ϕ_i denotes the scalar field (and its vacuum $|0_i\rangle_\phi$) for a black hole at location $\mathbf{x}_i$ for $i = 1, 2$.

We therefore find an interference term that is a genuine quantum effect extracted from the Unruh–DeWitt detector due to the black hole being in a quantum superposition.

9.4.2 Numerical analysis

In Fig. 9.4 we plot the measurement probabilities given in Eq. (9.37), rescaled by a normalization factor $\mathcal{N} = \sigma\lambda^2/2|\langle E|\mu|E_0\rangle|^2$ that depends on details of the detector's internal structure, as a function of the ratio R_2/R_1 of the radial coordinate distances between the detector and black hole in the branches of the superposition.

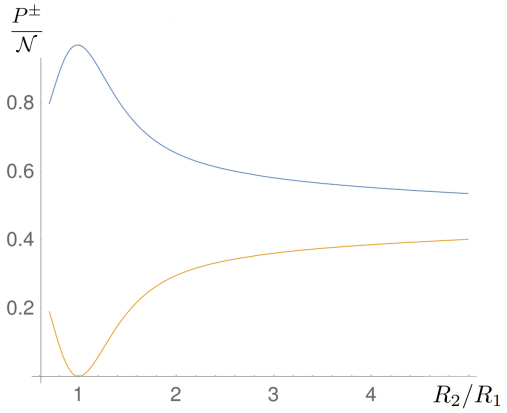

Figure 9.4: Measurement probabilities $P^\pm_E/\mathcal{N}$ with $\mathcal{N} = \sigma\lambda^2/2|\langle E|\mu|E_0\rangle|^2$ for a position-superposed black hole as a function of R_2/R_1. Here $P^+_E/\mathcal{N}$ is in blue and $P^-_E/\mathcal{N}$ in orange. Parameters used: $l/\sigma = 5, \sigma = 1$, $\tau_f = 10\sigma = -\tau_i$ in the detector's own clock time, and $R_1/\sigma = 2r_h, Ml^2 = 4, E - E_0 = 0.00016$.

We see here the effect of the interference term, which is proportional to the difference between the two graphs. The interference is peaked around $R_1 = R_2$, where the

(coordinate) distance between the black hole and detector is the same in the two branches, and decreases with increasing ratio and as the distances between the detector and black hole increase in both branches, an effect in line with the decay of the conformally coupled scalar field two-point function with distance, i.e. due to $W^{12}_{BTZ}(R_1, R_2)$ in Eq. (9.38). However, right at and very near to $R_1 = R_2$, this calculation cannot be trusted, as then the black hole is not in a superposition of locations (as seen from the detector). In this case, we have $|+\rangle_B = 2/\sqrt{2}\,|x_1\rangle\,, |-\rangle = 0$ meaning that the $|\pm\rangle$ measurement is ill-defined and the assumed initial state in Eq. (9.21) with $\mathbf{x}_1 = \mathbf{x}_2$ is not normalized there, such that the calculation of the interference probabilities in Eq. (9.37) cannot be trusted there.

Furthermore, it is meaningful to compare our results with those of Ref. [119], where a mass-superposed black hole was considered, instead of the position-superposed one we consider. In particular, we are comparing the measurement probabilities in Eq. (9.37), with the measurement probabilities of Eq. (11) of Ref. [119], that are analogous to the setup studied here, but when considering a black hole in superposition of masses. We plot their probabilities in Fig. 9.5, reproducing Fig. (1) of Ref. [119]. The authors argue that the sharp peaks that they observe are a consequence of having a superposition of different masses, and in particular interpret it in light of black hole mass quantization [716–720]. Our results further support this claim: when the black hole is in superposition of locations, but with single-valued mass, the sharp peaks are indeed absent. In the next subsection we will show in detail how this phenomenon arises from the spectrum probed by the detector.

9.4.3 Mass- vs. position-superposition

In this subsection we derive an analytical expression for the spectrum probed by the detector in the case of a mass- and position-superposed black hole. We use it to identify the origin of the peaks that were observed numerically in the detector response for a mass-superposed black hole in [119] (see Fig. 9.5). We find that the peaks can be traced to a coherent enhancement of the image sum at rational square-root mass ratios, and that additional singularities of the spectrum arise at certain rational mass ratios and corresponding position values in the mass-superposed case. Concretely, we will consider a generic mass- and position-superposed black hole, with position and mass R_1, M_1 and R_2, M_2 in the two branches of the superposition, respectively. Details can be found in Appendix B.6.

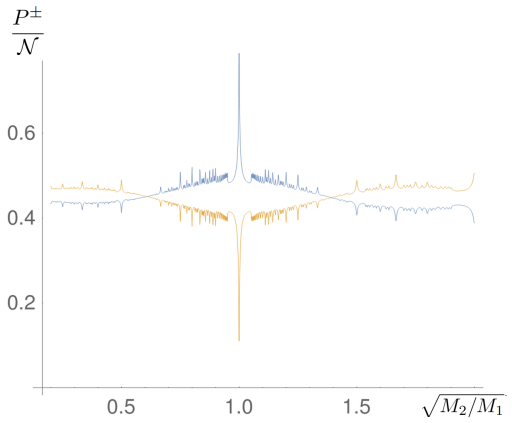

Figure 9.5: Reproduction of Fig. (1) of Ref. [119]: Measurement probabilities $P_E^{\pm}/\mathcal{N}$ with $\mathcal{N} = \sigma\lambda^2/2|\langle E|\mu|E_0\rangle|^2$ for a mass-superposed black hole as a function of $\sqrt{M_2/M_1}$. Here $P_E^{+}/\mathcal{N}$ is in blue and $P_E^{-}/\mathcal{N}$ in orange. Parameters used: $l/\sigma = 5, \sigma = 1$, $\tau_f = 5\sigma = -\tau_i, R/\sigma = 25, M_1 l^2 = 4, E - E_0 = 0.01 M_1 (= \Omega$ in Ref. [119]).

Single-spacetime detector response. In a non-superposed spacetime, the detector response from Eq. (9.11) can be expressed as [10]

$$\mathcal{F}(E) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, |\widehat{\eta}(\tfrac{\omega}{\gamma})|^2\, \widehat{W}_{BTZ}(E\gamma + \omega), \tag{9.40}$$

with time integration limits pushed to $\pm\infty$ and the Fourier transform defined as

$$\widehat{f}(\omega) = \int_{-\infty}^{\infty} ds f(s)\, \mathrm{e}^{-\mathrm{i}\omega s}. \tag{9.41}$$

An analytical expression for $\widehat{W}_{BTZ}(K)$ can be found [746], see Appendix B.6 for details.

Interference term in detector response. We can also express the interference term $\mathcal{F}_{12}(E)$, both for mass- and position-superposed black holes, in a similar form as Eq. (9.40), obtaining

$$\mathcal{F}_{12}(E) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{d}\omega\, \widehat{\eta}(\tfrac{\omega}{\widetilde{\gamma_1}})^*\, \widehat{\eta}\left(\tfrac{\omega}{\gamma_2} + E[\tfrac{\widetilde{\gamma_1}}{\gamma_2} - 1]\right) \widehat{W}_{BTZ}^{(12)}\left(\omega + \widetilde{\gamma_1} E\right). \tag{9.42}$$

Following Ref. [119], we consider AdS time coordinates $\bar{t}$ related to the BTZ coordinates t as $\bar{t} = t\sqrt{M}$ (see Supplemental Material of Ref. [119], for instance). Using the formulas in Appendix B of Ref. [746], an analytical expression for $\widehat{W}_{BTZ}^{(12)}$ can be found, with the additional subtlety that for $\sqrt{M_1/M_2} \in \mathbb{Q}\backslash\{1\}$ and $R_1/\sqrt{M_1} = R_2/\sqrt{M_2}$ an additional

[10] See Ref. [808] for a more precise treatment.

pole arises, giving a singular term sing.:

$$\begin{aligned}\widehat{W}^{(12)}_{BTZ}(K) = \text{sing.} + \frac{1}{2\sqrt{\widetilde{\gamma}_1\widetilde{\gamma}_2}}\frac{1}{\sum_k \Upsilon^{2k}}\frac{1}{e^{K/T_H^{AdS}}+1} \\ \sum_{m,n}\left[P_{\frac{iK}{2\pi T_H^{AdS}}-\frac{1}{2}}\left(\beta^{(12)}_{mn}\right) - \zeta P_{\frac{iK}{2\pi T_H^{AdS}}-\frac{1}{2}}\left(\beta'^{(12)}_{mn}\right)\right],\end{aligned} \tag{9.43}$$

with T_H^{AdS} the Hawking temperature of the AdS–Rindler horizon with respect to AdS–Rindler time, $T_H^{AdS} = 1/(2\pi l), \widetilde{\gamma}_i = \gamma(R_i, M_i)/\sqrt{M_i}$, and $\beta^{(12)}_{nm}, \beta'^{(12)}_{nm}$ factors that depend on M_1, M_2, R_1, R_2. Here P_ν is a Legendre function of the first kind. If $\sqrt{M_1/M_2} \notin \mathbb{Q}\backslash\{1\}$, then we have sing. $= 0$, else it gives the following contribution if also $R_1 = R_2\sqrt{M_1/M_2}$:

$$\begin{aligned}\text{sing.} =& \frac{1}{4\sqrt{\widetilde{\gamma}_1\widetilde{\gamma}_2}pq} \\ &\text{if } \sqrt{\frac{M_1}{M_2}} = \frac{p}{q} \in \mathbb{Q}\backslash\{1\} \text{ and } \frac{R_1}{R_2} = \sqrt{\frac{M_1}{M_2}}\end{aligned} \tag{9.44}$$

with $\gcd(p, q) = 1$. Here we used a natural cut-off procedure for the infinite sums $\sum_{m,n}$ as suggested by the group structures that define the BTZ spacetime through the periodic identification of AdS–Rindler space, see also Appendix B.6.

For other values of R_1, R_2, one must resort to the 'regular part' of $\widehat{W}^{(12)}_{BTZ}(K)$ of Equation (9.43). For rational $\sqrt{M_1/M_2} = \frac{p}{q}$ and R_2/R_1 not this same rational number but R_1, R_2 fixed, $\beta^{(12)}_{mn}$ has the same value for terms in the sum over m, n satisfying $m/n = \sqrt{M_1/M_2}$, and thus add up coherently. In addition, for R_1, R_2 both large compared to the horizon radius, $\beta^{(12)}_{mn} \to 1+$, the Legendre function $P_{-\frac{1}{2}-iK\ldots}$ remains close to 1 until large $|K|$-values before decaying leading to sharper peaks. On the other hand, for irrational $\sqrt{M_1/M_2}$, for large $|K|$ oscillatory behavior (decaying with $|K|$) of the Legendre function dephases the sum leading to cancellations on average. For near-rational irrational $\sqrt{M_1/M_2}$ cut-off dependent peaks may occur at large R_1, R_2, but only limitedly as determined by the irrational offset, with more averaged cancellations for large $|K|$. Additionally, large positive K-values are exponentially suppressed by the thermal factor, and thus main contributions to (peaked) behaviour come from large negative K-values.

This leads to the peaks in the interfered response function in Ref. [119], as pictured in Figure 9.5.

9.5 Discussion and outlook

We have analyzed the transition probability of a static detector in a position-superposed BTZ black hole spacetime, where our measurement loses which-way information on the position of the black hole. To do so, we employed a quantum reference frame transformation, showing that the physical situation is equivalent to a quantum-controlled position-superposed Unruh–DeWitt detector evolving in a classical spacetime. As a result, we find a 'nonclassical' term that would be absent for a black hole in a probabilistic mixture of locations.

The same setup but for a mass-superposed BTZ black hole was considered by Foo et al in Ref. [119], where the interferometric measurement was performed upon superposed mass states for the black hole, and sharp peaks reminiscent of black hole quantization were found. In our setup we find no such sharp peaks, providing further evidence that those peaks are a consequence of black hole mass quantization. Moreover, we have provided analytical expression for the spectrum probed by an Unruh–DeWitt detector in the case of a mass- and position-superposed BTZ black hole. These expressions show that the sharp peaks originate from a coherent enhancement of the image sum at rational square-root mass ratios, and that at certain position ratios an additional pole singularity appears at rational square-root mass ratios.

Building on our results, and more broadly on the study of detectors in superposed spacetimes, several promising avenues for future research emerge. A natural direction for future work is to investigate the physical regimes in which our approximations break down. For instance, we have assumed that the two branches of the superposition are classically perfectly distinguishable, so that the corresponding black hole states can be treated as orthogonal. This assumption is expected to fail when the two superposed locations become arbitrarily close. A first step toward addressing this regime could be to apply linearized quantum gravity, along the lines of Refs. [680, 681]. In addition, we might relax the assumption of gravity being sourced only by position. In the more general situation in which gravity is sourced by both position and momentum, a more refined treatment of semiclassical states is required, for example in terms of gravitational coherent states [102, 104, 443, 809–816].

Furthermore, another interesting possibility would be to investigate coherences in a detector with many energy levels near a superposed BTZ black hole, as was done for the Schwarzschild black hole in Ref. [655]. Moreover, we have investigated the first

nontrivial perturbative order in the Unruh–DeWitt interaction, but a more complete understanding of this interaction requires the investigation of higher-order terms [817, 818] and nonperturbative treatments [819, 820].

As the quantum BTZ black hole has an effective description as a two-dimensional conformal field theory (CFT) at infinity that can be coupled to classical matter [771, 821, 822], it would be interesting to investigate the results regarding transition amplitudes and probabilities of detectors in superposed BTZ black hole spacetimes in the related CFT. One would expect similar effects as for the position- and mass-superposed BTZ black holes in other superposed spacetimes, such as superposed Schwarzschild black holes for example, and how the spectra of potential peaks in solvable 1+1d (black hole) spacetimes [395] relate to CFTs on the boundary in AdS/CFT [74]. Our analytical results suggest that this could be achieved by probing the spectrum of interfered two-point functions between superposed spacetimes, and look for additional Hadamard-like singularities [126, 193, 194, 196] in these interfered two-point functions that relate properties such as black hole mass with the quantized angular structure of the quantized field solutions. In general, calculating Wightman functions in such superpositions might be challenging, but branch-dependent coordinate transformations, a form of a quantum reference frame transformation, or embeddings in higher-dimensional (Minkowski) spacetime may offer potential routes. An example of the latter is Schwarzschild in 6d Minkowski spacetime [823] in the GEMS (Global Embedding in Minkowski Spacetime) paradigm [824–829], even though it is not always clear when (in which cases) exactly this mapping works [830].

Finally, this chapter combines aspects of quantum reference frames, quantum gravity and detector models, and thus works toward an operational understanding of quantum theories of gravity. Such operational approach allows to critically assess some common assumptions of black hole paradoxes [46, 49, 388, 508], especially when combining the viewpoints of different observers such as the cloning [47, 48] and firewall paradoxes [49–52], as mentioned in Section 8.4.1. Explicit calculations using quantum reference frames and detectors probing the radiation can further quantify these subtleties. For example, claims concerning the entanglement of subsystems can be probed operationally with entanglement harvesting [649, 668–677, 748–751], including claims of early and late radiation entanglement, and which observers would detect high entanglement across the horizon and in the radiation from entangled black holes [831–833].

— 10 —

Conclusions and future work

10.1 Conclusions

In this thesis, we have discussed several paradoxes and setups that push quantum theory's applicability to its limits. For each part, we summarize our contributions and outline future directions.

10.2 Part I: discussion and outlook

In Part I, we have analyzed paradoxes that sharpen the tension within the measurement problem. The central assumption was the Universality of Quantum Theory, under which measurements performed in sealed laboratories can be modeled unitarily. In other words, we assumed quantum theory applies to all systems alike, leading to the scenario of Wigner's friend and extensions thereof. Our main contributions of Part I include a refined version of the Frauchiger–Renner paradox, connections between extended Wigner's friend scenarios and notions of nonclassicality, and an argument against the absoluteness of free choices. We briefly discuss the implications of these paradoxes by categorizing potential resolutions, after which we indicate directions for future work.

Categorizing resolutions to extended Wigner's friend paradoxes. Resolutions of extended Wigner's friend arguments fall into three broad categories.

a. Refuting Universality of Quantum Theory. A first category refutes the existence of superobservers, i.e., rejects the Universality of Quantum Theory. This would require further explanation of why quantum theory cannot be applied equally to all systems, and

why a measurement cannot be modeled unitarily. We have not probed this possibility further. See also the paragraph *Towards actual implementations of extended Wigner's friend paradoxes* below.

b. From Absoluteness to the Relativity of Observed Events. A second category refutes the Absoluteness of Observed Events, meaning outcomes of performed measurements cannot be absolute. This requires a replacement principle. As discussed in Chapter 3, our refined Frauchiger–Renner paradox suggests a natural resolution, the Relativity of Observed Events, paraphrased à la Peres's dictum as 'Unperformed measurements have no results, and unknown results have no values'. An appealing further direction is to explore whether quantum theory can be reformulated or reconstructed by taking the Relativity of Observed Events as a guiding principle, and connect with proposed relational reformulations [320, 325–327]. Our scenario combining Wigner's friend with the Pusey–Barrett–Rudolph theorem [38] of Chapter 5 moreover motivates to extend such relativity to the notion of free choices. As an extremely fundamental notion in physics, while at the same time deeply connected to foundational philosophical questions, rethinking the notion of free choice and 'interventions' in this light may prove a valuable task (see, e.g., Ref. [269] for work in this direction).

c. Fine-tuning resolutions. A third category for resolving extended Wigner's friend arguments is to hold on to the Absoluteness of Observed Events (AOE), at the cost of adopting a principle of fine-tuning, i.e., at the cost of a rather conspiratorial view of nature. In the case of the Local Friendliness no-go theorem, this involves refuting the assumption of Local Agency. We have further elaborated on this fine-tuning underlying the Local Friendliness assumption of Local Agency in Section 4.4 by outlining a no-go theorem generalizing the Local Friendliness no-go theorem. It would be interesting to connect in more detail to Ref. [33], who consider the Local Friendliness result in light of causal models and fine-tuning, as opposed to operational fine-tuning in our no-go result. On a technical level, the scenario employed in our result combined Wigner's friend with generalized noncontextuality, and still holds in background-independent theories, such as some theories of quantum gravity, as opposed to the Local Friendliness no-go theorem. Specifically, the Local Friendliness assumption of Local Agency can be seen as a generalization of no-signaling, and refuting it allows for some 'unobserved' faster-than-light signaling. The fine-tuning then lies in the fact that, even though unobserved signaling would be allowed, this signaling can never be observed.

In the Frauchiger–Renner paradox, and our refined version of it of Chapter 3, the fine-tuning resolution lies in rejecting the possibility to formulate correct predictions using

quantum theory for certain outcome values, even though those outcome values exist and are definite when assuming the Absoluteness of Observed Events. To explain this somewhat more, consider the following setup. Alice has performed a measurement, and Bob wants to formulate a prediction for her outcome value a. Assuming Absoluteness of Observed Events, Alice's outcome value exists and is absolute and single-valued, irrespective of anything else in the universe. Therefore, Alice's outcome a exists, and is single-valued, for Bob as well. Suppose that Bob uses quantum theory, and finds that Alice should obtain $a = 0$. The fine-tuned resolution then does not allow Bob to assert the prediction 'I know that Alice obtained $a = 0$'; in some cases, this prediction may be correct, but in other cases, such as when superobservers act as in the FR paradox, it may be wrong. How, then, should Bob formulate his prediction? One could argue that little room remains to do so, and one way would be to let Bob make the statement: 'I know that I would find that Alice has $a = 0$ if I ask for her outcome.' This, however, closely resembles the practical implementation of the Relativity of Observed Events, and one may question the practical benefit of assuming such hidden, fine-tuned additional layer of reality. Historically, simplicity is preferred, all else equal.

Nonclassicality and extended Wigner's friend arguments. More technically, we have provided several arguments combining extended Wigner's friend paradoxes with notions of nonclassicality. In fact, also the Pusey–Barrett–Rudolph theorem can be seen as capturing a form of nonclassicality or quantum advantage in some tasks [834]. This motivates further exploration of such arguments, and to probe what is nonclassical about quantum theory assuming its universality, and about Wigner's friend paradoxes [292]. Moreover, it would be interesting to consider generalizations of the extended Wigner's friend scenarios based on fine-tunings that go beyond those linked to contextuality [223, 224, 226, 228], such as violations of bounded ontological distinctness [262, 356].

On the Relativity of Observed Events and time symmetry. Above, in category *b. From Absoluteness to the Relativity of Observed Events*, we have argued that assuming the Universality of Quantum Theory and adhering to no fine-tuning, our refined FR paradox leads to the resolution of replacing the Absoluteness of Observed Events by the Relativity of Observed Events. A world governed by the Relativity of Observed Events rather than Absoluteness may be favored also on the basis of time symmetry for physical laws; under AOE, once an observer has performed a measurement, its outcome value is absolute and thus persists, even if its physical records are erased by a subsequent observer.

The only way in which one could save time symmetry would then be by assuming it was pre-existing at all times, leading to even further requirements on this hidden layer of reality. A possibility of probing this argument in more detail would be to combine Wigner's friend with the 'timelike Bell scenario' as was used to argue against no fine-tuned time-symmetric ontological models in Ref. [230]; the scenario of Section 4.4 and Ref. [22] may provide a starting point for this investigation. The view of the arrow of time as an emergent time asymmetry from time-symmetric physical laws [269, 835–839] is in line with a practical world with abundance of decoherence and practical Absoluteness of Observed Events despite adhering to the Relativity of Observed Events on a fundamental level.

Another consideration regarding spacetime is the role of gravity and spacetime in the decoherence of quantum systems [138, 301, 601, 602, 605, 645, 840–847]—not necessarily invoking modifications to quantum theory or clashing with the equivalence principle as argued in Refs. [310, 311], see, e.g., Ref. [645, 848]— and, potentially, matter-gravity entanglement in the thermodynamic arrow of time [849]. Although not necessarily a resolution to the measurement problem, these phenomena may contribute to the practically observed decoherence of quantum systems and the emergence of classicality in our world, and the resulting apparent absence of superobservers [697], see also the paragraph below.

On the reality of the wavefunction. As discussed in Section 2.2, studies of nonclassicality provide no-go results for a no fine-tuning interpretation of quantum theory within the ontological models framework of classical realism. Regarding the question whether the quantum wavefunction can be regarded as ontic or epistemic in this framework, the Pusey–Barrett–Rudolph theorem [38, 231] implies that it must be ontic, modulo some assumptions regarded as relatively weak. These nonclassicality results invite us to deepen our understanding of quantum theory beyond (classical) realism, with moreover extended Wigner's friend arguments suggesting a form of relationalism, even for free choices, as we have argued in Chapter 5 and category *b.* above. The Relativity of Observed Events is such a principle, and it would be interesting to probe what relational frameworks implementing this principle tell us about the nature of the wavefunction.

Towards actual implementations of extended Wigner's friend paradoxes and on the absence of superobservers. Arguably, a simple qubit would not qualify as an observer, at least not in a naive way [267]; a more complex system is required, while not losing its

coherence to an uncontrollable environment, i.e., an environment the superobserver does not have (quantum) control over. Perhaps the closest to an observer playing the role of a friend in future experiments is a quantum artificial intelligence, i.e., an artificial intelligent program run on a quantum computer, passing some Turing tests if one insists. We already have AIs; if we can build a sufficiently powerful quantum computer, implementing such a friend should not be a problem. An experimental effort of realizing the Local Friendliness scenario with quantum systems serving as the friends, defining a 'branch factor' to quantify the 'observerness' of the system, can be found in Ref. [850].

However, as discussed briefly in Section 8.4, we can ask ourselves, where are *our* superobservers? Why can we trust our memories? If Wigner can manipulate the brain of his friend, Wigner could implant memories and experiences that would lead the friend to different physical laws, or even nonsensical results. Why are we not a Boltzmann brain [292]? Does the concept of an observer require an emergent level of description that we cannot fully grasp yet? Computational complexity, quantum error correction, chaos and random matrices have been associated with black holes, holography and the decoding of Hawking radiation [49, 53, 474–477, 479, 480, 583, 686–696]; it would be interesting to see whether these concepts play a role as well in the emergence of classicality and defining observers more rigorously.

The question of the practical absence of superobservers [697] is closely related to the emergence of kinematic and dynamical classicality, which can be approached through the decoherence approach [6, 307, 698–702] or, relatedly, the approach of coarse-grained observables [5, 703–708]. In fact, the concepts of (classical) chaos and random matrices have appeared in the decoherence and coarse-graining approaches to the emergence of classicality [5, 709, 710], and topics such as computational complexity, chaos, quantum error correction and scrambling in measurement-induced phase transitions [711–715]; it would be interesting to further probe connections between these areas.

Connecting to Gödel? There are certainly analogies between Gödel's incompleteness theorem and the liar cycles that arise in extended Wigner's friend paradoxes (and quantum theory's Kochen–Specker contextuality they are based upon [275]). Prior work has explored this connection in some more detail [851, 852], but it remains to be seen whether this connection is more than a mere analogy [853].

What if quantum theory is not the final story? Finally, it is, of course, possible that quantum theory is not the ultimate description of nature, a proposal being a 'gravitized

quantum theory', for example [854]; if this is the case and such theory has nontrivial implications for the definition of observers, perhaps contrary to current expectations, it might have nontrivial repercussions for Wigner's friend paradoxes.

10.3 Part II: discussion and outlook

In Part II, we have discussed some black hole paradoxes and considered a setup where quantum aspects of gravity are combined with quantum field theory. In this part, we assumed that quantum theory also applies to gravity, another instance of assuming its universal validity. Our main contributions of Part II consist of combining Wigner's friend and black hole physics into a unified argument that strengthens existing black hole paradoxes, a critical analysis of standard assumptions therein, and a calculation of a detector's response near a superposed black hole.

In Chapter 7, after briefly outlining some prominent black hole paradoxes, we have highlighted the many extra degrees of freedom in the form of soft hair that were first overlooked by Hawking when formulating his information loss argument for black holes—recognized by Hawking himself about fifty years later [101]. Together with the assumption that quantum theory applies to gravity as well, a first argument by conservation laws suggests the black hole to become highly superposed, and provides a possibility for information recovery and information transfer to the radiation.

While providing a possibility towards resolving the black hole information puzzle, this does not yet resolve black hole paradoxes featuring infalling observers such as the cloning and firewall paradoxes [47–52, 393], which was the topic of Chapter 8. In this chapter, after formulating scenarios that strengthen these paradoxes by combining Wigner's friend and black hole physics into a unified argument, going beyond the analogy exposed by Hausmann and Renner [107], we turned to a critical assessment of the underlying black hole assumptions. These assumptions are inherited from the standard formulations of the black hole cloning and firewall paradoxes. We have highlighted several important subtleties regarding these assumptions in Section 8.4, related to superposed spacetimes, gauge invariance and locality[1] concerning information recovery and observables for infalling

[1] By locality we mean here locality in the quantum theory itself, i.e., how to localize systems and events in quantum field theory and quantum gravity, and not the classical locality or classical local causality as in Bell's theorem.

observers. These nontrivialities are not problems of the formalism or setup used, but are crucial *features* of any theory of quantum gravity.

Among the methods that can be employed for further probing these features and their implications for black hole paradoxes are quantum reference frames and detectors, or, more generally, a shift to focusing on operational questions, tied to the experience of specific observers. We have presented a calculation combining quantum reference frames and a detector in Chapter 9, by analyzing the response of an Unruh–DeWitt detector near a position-superposed BTZ black hole, and identifying an 'interference' term absent in classical gravity. To obtain the interaction of a detector with the field on a superposed spacetime, we have used a quantum reference frame transformation to the case of a superposed detector near a black hole with definite position, which yields a non-superposed spacetime.

The choice of a position-superposed 2+1d BTZ black hole, instead of other possible superposed spacetimes, was motivated by the study of a *mass-superposed* black hole in Ref. [119], where peaks in the detector response as a function of the ratio of the square root of the masses were found, argued to be reminiscent of black hole mass quantization. The absence of peaks found in our response for a *position-superposed* black hole further strengthens this claim. We have also provided an analytical expression for the spectrum probed by the detector, with peaks in the mass-superposed case corresponding to singularities that would give the Hadamard singularity structure in an otherwise classical two-point function. Specifically, these singularities arise from an interplay of the black hole mass with a discrete image sum, which can be seen as arising from a 'winding number' for functions of the angular coordinate on the 1-dimensional circle S^1 (as the BTZ black hole is 2+1-dimensional). This connection arises in the BTZ case of course from the definition of the BTZ black hole as a quotient of AdS that defines its mass parameter. In the case of a Schwarzschild black hole, this points toward searching for singularities in superposed black hole two-point functions arising from an interplay between the black hole mass and spherical harmonic labels l, m in terms leading otherwise to Hadamard singularities in non-superposed black hole two-point functions.

Whereas the calculation of Chapter 9 exemplifies an operational setup of a detector and quantum reference frames, it marks only a first step in the unfolding story of this operational approach. Further development of these methods in physical setups concerning cosmology and black hole physics is required. As argued above, and more generally in Part II of this thesis, its application may yield interesting insights into the

physics of black holes and its paradoxes.

— A —

Extra black hole protocols of Chapter 8

A.1 A Local Friendliness version of the first protocol

The protocol is similar to our first protocol, but now Bob additionally makes a choice whether to reveal his outcome to Wigner or not before falling into the black hole, and Alice encodes her outcome a in the computational Z-basis of a qubit S_C which she sends to another observer Charlie before falling in the black hole. Charlie makes a choice $x = 0, 1$ whether to measure S_C in the Z-basis (decoding Alice's outcome) for $x = 0$ or in the $|\pm\rangle$ basis for $x = 1$.

Theorem 10. *If a superobserver can decode the Hawking radiation and reconstruct the quantum state of an infalling observer, and still receive a short message from that observer sent inside the black hole and assumed to be unaltered by actions outside, an operational violation of no-signalling is found.*

Proof of Theorem 10. Consider the following protocol.

- Alice and Bob, residing in sealed labs, share a system $S_A S_B$ in the entangled Hardy state of Eq. (8.1), and each measure their qubit in the computational basis, modeled unitarily as U_A, U_B.
- Alice encodes her outcome in the computational basis of a qubit S_C, which she sends to Charlie, and next Alice falls into the black hole.
- Charlie has a choice $x = 0, 1$ whether to measure the qubit in the computational Z-basis or in the $|\pm\rangle$ basis.
- Somewhat later, Bob makes a choice $y = 0, 1$ to reveal his outcome to Wigner or not, and falls into the black hole.

- Alice and Bob, once inside the black hole, try to send their outcomes to Ursula and Wigner (who will fall into the black hole somewhat later).
- Before falling in, Ursula and Wigner retrieve the Hawking radiation containing the states of Alice's and Bob's labs, apply the 'undoing operation' $U_A^\dagger, U_B^\dagger$, respectively, to the decoded lab states of Alice and Bob in the Hawking radiation and measure S_A, S_B, respectively, in the $|\pm\rangle$ basis.
- Next, Ursula, Wigner and Charlie all fall into the black hole, aiming to still obtain the outcomes a, b of Alice and Bob.

Regarding the spacetime arrangement (important as our argument will involve no-signalling), Alice falls in the black hole before Bob, such that Ursula can retrieve Alice's Hawking radiation outside of the future lightcone of Bob's choice y, and obtain u, b outside of the future lightcone of y. Moreover, Charlie makes his measurement choice x such that Wigner's measurement on Bob's Hawking radiation takes place outside of the future lightcone of x. The spacetime arrangement is sketched in Figure A.1.

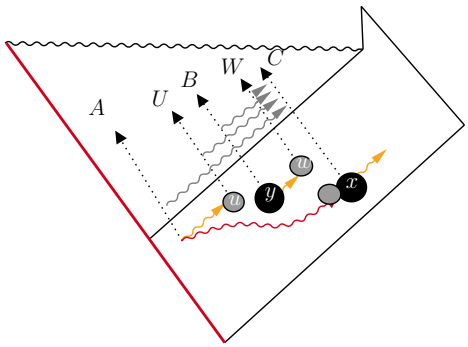

Figure A.1: The LF-black hole protocol in a Penrose diagram; x, y denote choices by Charlie and Bob, respectively, and the smaller black dots denote operations by Alice, Ursula, Bob, Wigner and Charlie.

Alice and Bob measure S_A, S_B, which can be modeled as unitaries U_A, U_B, and, more specifically, as $CNOT$ gates if we assume their initial memory states are $|0\rangle_A, |0\rangle_B$, i.e.

$$U_A = CNOT_{S_A A} = \Pi_0^{S_A} \otimes \mathbb{1}_A + \Pi_1^{S_A} \otimes (|1\rangle\langle 0| + |0\rangle\langle 1|)_A, \tag{A.1}$$

with $\Pi_0 = |0\rangle\langle 0|$, and similarly for Bob's measurement and Alice's action of preparing her outcome in the qubit S_C she sends to Charlie. Thus, the actions described thus far are

$$CNOT_{S_A S_C} CNOT_{S_A A} CNOT_{S_B B} |\psi_0\rangle_{S_A S_B} |000\rangle_{A B S_C}, \tag{A.2}$$

where we are sloppy about the tensor product ordering in the initial state (we omit some tensor products $\otimes$ and swap operations to bring the initial state into the ordering of the operations that act on it, but this should be clear from the use of indices).

Note that when Charlie obtains the outcome $c = +$ for $x = 1$, we have that

$$|+\rangle \langle +|_{S_C} CNOT_{S_A S_C} = |+\rangle \langle +|_{S_C} \otimes \mathbb{1}_{S_A}. \tag{A.3}$$

Therefore, if Charlie obtains $c = +$, then Ursula's undoing of Alice's measurement will indeed reset her lab back to the initial state, the same that would be obtained if Alice did not prepare her outcome in the computational basis of the qubit S_C.

Assuming all involved outcomes a, b, u, w, c can be brought together inside the black hole, the argument goes as follows.

First of all, as one can verify using quantum theory's predictions and Eq. (A.3), we have

$$p(u{=}-, w{=}- \,|x{=}1, y{=}1, c{=}+) > 0. \tag{A.4}$$

Next, quantum theory predicts that

$$p(u{=}-, b{=}0|x{=}1, y{=}0, c{=}+) = 0.$$

As all of u, b, c are obtained spacelike separated from each other and not in the future lightcone of y, we can give a no-signalling argument that the joint outcomes u, b, c should be independent of y. Namely, we could give Ursula and Bob extra choices $x_u, x_b \in \{0, 1\}$ whether (for choice 1) or not to encode their outcomes in the computational basis of extra qubits that they send to Charlie. Then by measuring these qubits in the computational basis, Charlie can obtain the correlations u, b, c if Ursula and Bob choose $x_u = 1 = x_b$. However, these correlations should be independent of x_u, x_b, y by no-signalling as these choices are all spacelike separated from where Charlie obtains these correlations ubc. Therefore, we obtain that

$$p(u{=}-, b{=}0|x{=}1, y{=}1, c{=}+) = 0.$$

Using quantum theory and no-signalling we have that

$$p(a{=}0, b{=}1|x{=}1, y{=}1) = p(a{=}0, b{=}1|x{=}0, y{=}0) = 0,$$

and thus also $p(a{=}0, b{=}1|x{=}1, y{=}1, c{=}+) = 0$.

Finally,

$$p(a{=}1, w{=}- \,|x{=}1, y{=}1) = p(a{=}1, w{=}- \,|x{=}0, y{=}1)$$

by no-signalling, and quantum theory gives that

$$\begin{aligned} &p(a{=}1, w{=}-\,|x{=}0, y{=}1) \\ &\quad = p(a{=}1, c{=}a, w{=}-\,|x{=}0, y{=}1) = 0, \end{aligned}$$

meaning that

$$p(a{=}1, w{=}-\,|x{=}1, y{=}1) = 0$$

and thus also $p(a{=}1, w{=}-\,|x{=}1, y{=}1, c{=}+) = 0$.

Thus, we obtain that the distribution $p(a, b, u, w|x{=}1, y{=}1, c{=}+)$ should satisfy the conditions that

$$\begin{aligned} &p(u{=}-, w{=}-\,|x{=}1, y{=}1, c{=}+) > 0, \\ &p(u{=}-, b{=}0|x{=}1, y{=}1, c{=}+) = 0, \\ &p(a{=}0, b{=}1|x{=}1, y{=}1, c{=}+) = 0 \\ &p(a{=}1, w{=}-\,|x{=}1, y{=}1, c{=}+) = 0. \end{aligned} \tag{A.5}$$

However, this cannot be true as then $u = -, w = -$ can occur, but $u = -$ implies that $b = 1$ and $w = -$ implies that $a = 0$, but $a = 0, b = 1$ cannot occur. ■

Comparison to the LF protocol. Note that, compared to the LF protocol [12], which involves Alice, Bob, Ursula and Wigner, we have an additional party involving Charlie who gets a qubit from Alice. This is because we cannot simply let Ursula and Wigner perform their measurements spacelike separated from the choices y, x, respectively, and dive in to obtain u, b jointly outside of the future lightcone of y and a, w jointly outside of the future lightcone of x. Namely, after Alice and Bob have performed their measurements, choose whether or not to reveal their measurements and fall in the black hole, Ursula and Wigner have to wait some time for the Hawking radiation containing the information about the lab states of Alice and Bob to come out of the black hole. Estimates of this timescale that Ursula and Wigner would have to wait for an old black hole are of the order of the scrambling time $r_s \log r_s$ [48] (with r_s the Schwarzschild radius), and the furthest we could position Ursula and Wigner away from each other are at antipodal locations from the black hole. But then light rays sent at the locations where Alice and Bob make their choices x and y might reach Wigner or Ursula, respectively, before they can decode the Hawking radiation, so that signalling is possible. The trick with having this additional prepare-and-measure part is the same one as applied in Chapter 4 to obtain EWF arguments based on contextuality rather than nonlocality.

A.2 No-signalling firewall argument

Consider the following protocol, a slightly altered firewall paradox. Alice holds the early radiation, and distills a qubit (possibly precomputed using Oppenheim and Unruh's argument if its computational complexity would require so [480]), maximally entangled with the (late-radiation) mode B that Bob holds near the horizon of an old black hole: $1/\sqrt{2}(|00\rangle + |11\rangle)_{AB}$. The mode B is also maximally entangled with Charlie's mode inside the black hole: $1/\sqrt{2}(|01\rangle - |10\rangle)_{CB}$. Bob chooses $y = 0, 1$ to measure his qubit B in the Z or X basis. Alice always measures her qubit A in the Z basis. Charlie chooses $x = 0, 1$ to measure his qubit in a basis such that Charlie and Bob violate Bell nonlocality, for example by choosing $-(X + Z)/\sqrt{2}$- and $(-Z + X)/\sqrt{2}$-measurements, such that the empirical probabilities $p(c_x, b_y|x, y)$ violate Bell nonlocality [34, 36] and thus by Fine's theorem [42, 247, 342] a global distribution $p(c_0, c_1, b_0, b_1)$ reproducing the empirical correlations $p(c_x, b_y|x, y)$ cannot exist. Now, quantum theory predicts that Alice and Bob will always have the same outcome $a = b$ when $y = 0$, as $\langle b, \neg b|_{AB}(|00\rangle + |11\rangle) = 0$ for $b = 0, 1$. Therefore, we have that $p(a, b, c|x, y = 0) = \delta_{a,b}p(b, c|x, y = 0)$ and thus also

$$\begin{aligned} p(c, b|x, y = 0) &= p(c, b = a|x, y = 0) \\ &= \delta_{a,b}p(c, a|x, y = 0) \\ &= \delta_{a,b}p(c, a|x, y = 1), \end{aligned} \tag{A.6}$$

where we used no-signalling to conclude that $p(c, a|x, y) = p(c, a|x)$; we can let Alice and Charlie perform their measurement before Bob's choice so that they obtain a, c spacelike separated from Bob's choice y. This implies that the empirical correlations of $p(c, b|x, y = 0)$ are those of $\delta_{a,b}p(c, a|x, y = 1)$. Therefore, the joint correlations $p(c, b|x, y = 0), p(c, b|x, y = 1)$ are equal and should both be marginals from the global distribution $p(a, b, c|x, y = 1)$, which is empirically accessible if Alice, Bob and Charlie can meet inside the black hole. If such a distribution $p(a, b, c|x, y = 1)$ exists, then by a gluing procedure [17, 21, 342] we can produce a global distribution reproducing all empirical correlations $p(c, a|x = 1, y = 1), p(c, a|x = 0, y = 1), p(c, b|x = 1, y = 1), p(c, b|x = 0, y = 1)$, which by Eq. (A.6) are the same empirical correlations as the empirical Bell scenario correlations $p(c, b|x, y)$ for $x, y \in \{0, 1\}$ that violate Bell nonlocality with $c, a|x, y = 1$ replaced by $c, b|x, y = 0$. However, as we explained a few lines ago, by Fine's theorem such global distribution cannot exist, and we find a contradiction.

The gluing procedure unfolds as follows. Denoting $c|x$ by c_x, for example using c_0 for the outcome variable c when $x = 0$, the correlations $p(c_0, b, a|x = 0, y = 1)$ can be written as

$$\begin{aligned} p(c_0,b,a|x=0,y=1) &= p(a,b|x=0,y=1)p(c_0|x=0,y=1,b,a) \\ &= p(a,b|x=1,y=1)p(c_0|x=0,y=1,b,a), \end{aligned} \tag{A.7}$$

using no-signalling for $p(a, b|x = 0, y = 1) = p(a, b|x = 1, y = 1)$, and noting that $p(c_0, b, a|x = 0, y = 1)$ is also empirically accessible if an observer as Charlie can obtain all of c_0, b, a inside the black hole. We also have

$$p(c_0,b,a|x=0,y=1) = p(a,b|x=1,y=1)p(c_1|x=1,y=1,b,a), \tag{A.8}$$

Therefore, the empirical correlations $p(c_0, a|x = 0, y = 1), p(c_0, b|x = 0, y = 1), p(c_1, a|x = 1, y = 1), p(c_1, b|x = 1, y = 1)$ can all be obtained from the "glued" global distribution

$$p(c_0, c_1, a, b) := p(c_1, a, b|x=1, y=1)p(c_0|a, b, x=0, y=1). \tag{A.9}$$

We thus obtain the following result.

Theorem 11 (No-signalling firewall argument)**.** *Assuming (Fi)-(Fiii) of Theorem 8, an operational violation of no-signalling is found if Alice, Bob and Charlie can meet inside the black hole.*

The meaning of no-signalling is determined by the semiclassical causal structure. By an 'operational' violation of no-signalling it is meant that this violation is experimental, witnessed by a single observer, i.e. that all relevant variables are jointly accessible (by Alice, Bob and Charlie if they can meet up inside the black hole). If Alice, Bob and Charlie cannot meet up inside the black hole, one could argue the violation is one of Local Agency [13] rather than no-signalling.

— B —

Appendix for Chapter 9

B.1 The BTZ black hole as a quotient of AdS_3

As we will later in this Appendix refer to the relation between AdS_3 coordinates and the BTZ black hole, we shortly review how the BTZ black hole is obtained as a quotient of AdS_3. We follow Ref. [119], nearly verbatim.

Three dimensional anti-de Sitter AdS_3 can be obtained as the submanifold defined by

$$-l^2 = -T_1^2 - T_2^2 + X_1^2 + X_2^2 \tag{B.1}$$

in the 2+2-dimensional flat space $\mathbb{R}^{2,2}$ with metric

$$\mathrm{d}s^2 = -\mathrm{d}T_1^2 - \mathrm{d}T_2^2 + \mathrm{d}X_1^2 + \mathrm{d}X_2^2. \tag{B.2}$$

Taking the coordinates $\bar{t}, \bar{r}, \bar{\phi}$ on this submanifold, defined by

$$\begin{aligned} T_1 &= l\sqrt{\frac{\bar{r}^2}{l^2}}\cosh\bar{\phi} \\ X_1 &= l\sqrt{\frac{\bar{r}^2}{l^2}}\sinh\bar{\phi} \\ T_2 &= l\sqrt{\frac{\bar{r}^2}{l^2}-1}\sinh\frac{\bar{t}}{l} \\ X_2 &= l\sqrt{\frac{\bar{r}^2}{l^2}-1}\cosh\frac{\bar{t}}{l} \end{aligned} \tag{B.3}$$

with $l > 0$ and ranges $-\infty < \bar{t} < +\infty, l < \bar{r} < +\infty, -\infty < \bar{\phi} < +\infty$, one obtains the AdS–Rindler metric

$$\mathrm{d}s^2 = -\left(\frac{\bar{r}^2}{l^2}-1\right)\mathrm{d}\bar{t}^2 + \left(\frac{\bar{r}^2}{l^2}-1\right)^{-1}\mathrm{d}\bar{r}^2 + \bar{r}^2\,\mathrm{d}\bar{\phi}^2. \tag{B.4}$$

Note that $\overline{\phi}$ is not an angular coordinate in the AdS–Rindler space. Defining new coordinates r, t, ϕ by $r = \overline{r}\sqrt{M}, t = \overline{t}/\sqrt{M}, \phi = \overline{\phi}/\sqrt{M}$ for some $M > 0$, and periodically identifying ϕ by taking the quotient under $\Gamma : \phi \to \phi + 2\pi$, one obtains the BTZ metric

$$\mathrm{d}s^2 = -\left(\frac{r^2}{l^2} - M\right)\mathrm{d}t^2 + \left(\frac{r^2}{l^2} - M\right)^{-1}\mathrm{d}r^2 + r^2\,\mathrm{d}\phi^2, \tag{B.5}$$

with now ϕ being an angular coordinate.

B.2 BTZ correlation functions

The correlation functions in the BTZ black hole can be obtained from the correlation functions of AdS_3 using the method of images [109, 855], as the BTZ solution is a quotient space thereof. We use the results of Refs. [109, 119, 653]. The correlation functions in the scalar vacuum of AdS_3 are given by

$$W_{\mathrm{AdS}}(x, x') = \frac{1}{4\pi l\sqrt{2}}\left[\frac{1}{\sqrt{\rho(x, x')}} - \frac{\zeta}{\sqrt{\rho(x, x') + 2}}\right], \tag{B.6}$$

with ρ the squared geodesic distance in the 2+2 embedding space of AdS_3:

$$\rho(x, x') = \frac{1}{2l^2}\left[(X_1 - X_1')^2 - (T_1 - T_1')^2 + (X_2 - X_2')^2 - (T_2 - T_2')^2\right]. \tag{B.7}$$

Furthermore, $\zeta \in \{-1, 0, 1\}$ specifies the boundary conditions at asymptotic infinity: Neumann, transparent or Dirichlet [746, 763, 776]. We take transparent boundary conditions $\zeta = 0$ for now, calculations for other boundary conditions are analogous. The effect of transparent boundary conditions is to "recirculate the energy, angular momentum, etc lost to timelike infinity, resulting in a well-defined (but rather unusual) conservation law" [777]. The BTZ correlation function is then given by

$$\begin{aligned} W_{\mathrm{BTZ}}(x, x') &= \frac{1}{\sum_n \Upsilon^{2n}} \sum_n \sum_m \Upsilon^n \Upsilon^m W_{\mathrm{AdS}}(\Gamma_M^n x, \Gamma_M^m x') \\ &= \sum_m \Upsilon^m W_{\mathrm{AdS}}(x, \Gamma_M^m x'), \end{aligned} \tag{B.8}$$

with Γ_M the identification action $\phi \to \phi + 2\pi\sqrt{M}$ and $\Upsilon = +1, -1$ for untwisted and twisted fields; we will take $\Upsilon = 1$. The Wightman functions are actually distributions,

and are given by taking the limit $\epsilon \to 0+$ of Eq. (B.8) with ρ replaced by [119, 573]

$$\rho_\epsilon\left(x, \Gamma^n x'\right) = \frac{rr'}{r_h^2}\cosh\left[\frac{r_h}{l}(\Delta\varphi - 2\pi n)\right] - 1 - \frac{\sqrt{\left(r^2 - r_h^2\right)\left(r'^2 - r_h^2\right)}}{r_h^2}\cosh\left(\frac{r_h}{l^2}\Delta t - \mathrm{i}\epsilon\right) \tag{B.9}$$

with $r_h = l\sqrt{M}$, with $\Delta\varphi, \Delta t$ the differences in angular and time coordinates of x, x' whose radial components are denoted by r, r' (with coordinates giving the metric in Eq. (9.1)).

Using this expression for ρ in the BTZ correlation function Eq. (B.8), we obtain

$$\begin{aligned} W_{BTZ}(x_1, x_2) &= \frac{1}{4\pi l\sqrt{2}}\sum_m \frac{1}{\sqrt{\frac{R_1R_2}{r_h^2}\cosh\left[(\theta-2\pi m)\frac{r_h}{l}\right]-1-\sqrt{\left(\frac{R_1^2}{r_h^2}-1\right)\left(\frac{R_2^2}{r_h^2}-1\right)}\cosh\left(\frac{r_h}{l^2}\Delta t-\mathrm{i}\epsilon\right)}} \\ &= \frac{1}{4\pi l\sqrt{2}}\frac{1}{\sqrt[4]{\left(\frac{R_1^2}{r_h^2}-1\right)\left(\frac{R_2^2}{r_h^2}-1\right)}}\sum_m \frac{1}{\sqrt{\alpha_m^{(12)} - \cosh\left(\frac{r_h}{l^2}\Delta t - \mathrm{i}\epsilon\right)}} \\ &=: W_{BTZ}^{12}(t - t'), \end{aligned} \tag{B.10}$$

with

$$\alpha_m^{(12)} = \frac{\frac{R_1R_2}{r_h^2}\cosh\left[(\theta - 2\pi m)\frac{r_h}{l}\right] - 1}{\sqrt{\left(\frac{R_1^2}{r_h^2} - 1\right)\left(\frac{R_2^2}{r_h^2} - 1\right)}}. \tag{B.11}$$

Recall that, unless stated otherwise, we are working under transparent boundary conditions, see Eq. (9.5); for Neumann and Dirichlet boundary conditions we will have an additional term, giving

$$\begin{aligned} W_{BTZ}(x_1, x_2) =& \frac{1}{4\pi l\sqrt{2}}\frac{1}{\sqrt[4]{\left(\frac{R_1^2}{r_h^2}-1\right)\left(\frac{R_2^2}{r_h^2}-1\right)}}\sum_m \left(\frac{1}{\sqrt{\alpha_m^{(12)} - \cosh\left(\frac{r_h}{l^2}\Delta t - \mathrm{i}\epsilon\right)}}\right. \\ &\left. - \frac{\zeta}{\sqrt{\alpha_m^{(\prime 12)} - \cosh\left(\frac{r_h}{l^2}\Delta t - \mathrm{i}\epsilon\right)}}\right) \end{aligned} \tag{B.12}$$

with

$$\alpha_m^{\prime(12)} = \frac{\frac{R_1R_2}{r_h^2}\cosh\left[(\theta - 2\pi m)\frac{r_h}{l}\right] + 1}{\sqrt{\left(\frac{R_1^2}{r_h^2} - 1\right)\left(\frac{R_2^2}{r_h^2} - 1\right)}}. \tag{B.13}$$

B.3 Interaction picture calculation

In this section we shortly sketch the calculation for the detector's transition amplitudes Eqs. (9.10), (9.36) and (9.37) in the interaction picture. The interaction picture (see for example [130, 197, 856, 857]) is an intermediate picture between the Schrödinger representation, where states carry all time dependence but fields are time independent, and the Heisenberg representation, where operators carry all time dependence but states are time independent.

More specifically, let the total Hamiltonian of the system be

$$\hat{H} = \hat{H}_{0,S} + \hat{H}_{\text{int}} \tag{B.14}$$

where $\hat{H}_{\text{int}}$ is the interaction Hamiltonian between the detector and quantum fields in the Schrödinger picture; $\hat{H}_{0,S}$ is the free Hamiltonian in the Schrödinger picture. In the interaction picture, operators evolve according to the free time evolution (without interaction, given by Hamiltonian $\hat{H}_{0,S}$) and states evolve according to the freely-time-evolved interaction Hamiltonian ${}_I\hat{H}_{\text{int}}(t) = e^{i\hat{H}_{0,S}t}\hat{H}_{\text{int}}e^{-i\hat{H}_{0,S}t}$. Here the pre-subscript I denotes the subscript for 'interaction picture'. For simplicity of notation we will denote ${}_I\hat{H}_{\text{int}}(t)$ simply by $\hat{H}_{\text{int}}(t)$ from now on as we will perform our calculations only in the interaction picture. States in the Schrödinger and interaction pictures are related by $|\psi(t)\rangle_I = e^{i\hat{H}_{0,S}t}\,|\psi(t)\rangle_S$. Thus, in the interaction picture operators solve the equations of motion with respect to $\hat{H}_0$, and states evolve according to

$$i\frac{d}{dt}|\psi(t)\rangle_I = \hat{H}_{\text{int}}(t)\,|\psi(t)\rangle_I \tag{B.15}$$

for time slicing t defining the states $|\psi(t)\rangle_I$. Defining a unitary operator $U(t,t_0)$ such that

$$|\psi(t)\rangle_I = \hat{U}(t,t_0)\,|\psi(t_0)\rangle_I\,, \tag{B.16}$$

this unitary must satisfy

$$\hat{H}_{\text{int}}(t)\hat{U}(t,t_0) = i\frac{d}{dt}\hat{U}(t,t_0) \tag{B.17}$$

with initial condition $\hat{U}(t_0,t_0) = 1$. Simply exponentiating does not work because $\hat{H}_{\text{int}}$ is an operator and thus there is an ordering ambiguity. A formal solution of this equation can be written as the Dyson series

$$U(t,t_0) = Te^{-i\int_{t_0}^{t}\hat{H}_{\text{int}}(t')dt'} \tag{B.18}$$

with T time ordering, which to second order in a perturbation expansion in the interaction is given by

$$\hat{U}(t_f, t_i) = \mathbb{1} - i \int_{t_i}^{t_f} \mathrm{d}t \hat{H}_{\text{int}}(t) - \int_{t_i}^{t_f} \mathrm{d}t \int_{t_i}^{t} \mathrm{d}t' \hat{H}_{\text{int}}(t) \hat{H}_{\text{int}}(t') + \mathcal{O}\left(\lambda^3\right) \tag{B.19}$$

One can also think of the Dyson series as iteratively using the integral equation

$$\hat{U}(t_f, t_i) = \mathbb{1} - i \int_{t_i}^{t_f} dt' \hat{H}_{\text{int}}(t') \hat{U}(t', t_i) \,. \tag{B.20}$$

In our case, the free Hamiltonian governs the evolution of the (free) field ϕ and detector monopole moment $\hat{\mu}$, and the interaction Hamiltonian (with respect to proper time τ) consists of a coupling between the detector's monopole moment and the field along the detector's path. We take the detector's path to be a classical variable here, and not a quantum operator. Below we will consider a detector in a background where a black hole is in a superposed location in a frame where the detector is at a single fixed location. If we were to move to the frame where the black hole is in a single location and the detector is in a superposed position, we would have to make the detector location an operator here, see also Section 9.3.2. Therefore, the interaction Hamiltonian is given by

$$\hat{H}^{\tau}_{int} = \lambda \hat{\mu} \otimes \hat{\phi}(\mathbf{x}) \tag{B.21}$$

in the Schrödinger picture, with $\mathbf{x}$ the position of the detector, so in the interaction picture we have

$$\hat{H}^{\tau}_{int}(\tau) = \lambda \hat{\mu}(\tau) \otimes \hat{\phi}(x(\tau)). \tag{B.22}$$

Let the total initial state be $|\psi(\tau_i)\rangle_S = |0\rangle_\phi |E_0\rangle_\mu$ of the scalar field vacuum and detector ground state (the detector's internal degree of freedom), and denote the final total state by $|\psi(\tau_f)\rangle_S$ in the Schrödinger picture. We take the detector to be held at a fixed spatial location rather than freely falling, as explained in Sections 9.3.1 and 9.3.2. These states carry full time evolution dependence, being states in the Schrödinger picture. They are related to the states in the interaction picture by

$$|\psi(\tau_i)\rangle_I = e^{i\hat{H}^{\tau}_{0,S}\tau_i} |\psi(\tau_i)\rangle_S \,, \tag{B.23}$$

so that using Eq. (B.16) we have

$$|\psi(\tau_f)\rangle_S = e^{-i\hat{H}^{\tau}_{0,S}\tau_f} \hat{U}(\tau_f, \tau_i) e^{i\hat{H}^{\tau}_{0,S}\tau_i} |\psi(\tau_i)\rangle_S \,. \tag{B.24}$$

To first order in $\hat{H}_{\text{int}}$ we then have

$$|\psi(\tau_f)\rangle_S = e^{-i\hat{H}^{\tau}_{0,S}\tau_f} e^{i\hat{H}^{\tau}_{0,S}\tau_i} |\psi(\tau_i)\rangle_S - e^{-i\hat{H}^{\tau}_{0,S}\tau_f} i \int_{\tau_i}^{\tau_f} \mathrm{d}\tau \hat{H}_{\text{int}}(\tau) e^{i\hat{H}^{\tau}_{0,S}\tau_i} |\psi(\tau_i)\rangle_S \,. \tag{B.25}$$

Then, writing out H_{int} and post-selecting on the detector making a transition to energy level E we obtain at time τ_f

$$i\lambda \langle E| \int_{\tau_i}^{\tau_f} \mathrm{d}\tau \mu(\tau) \hat{\phi}[x(\tau)] |0, E_0\rangle \,, \tag{B.26}$$

where coordinate times t_i, t_f are related to the detector's proper time by $t_i = \tau_i/\gamma_J, t_f = \tau_f/\gamma_J$ with redshift factor γ_J given in Eq. (9.2). As time evolution for $\hat{\mu}(\tau)$ is given by

$$\hat{\mu}(\tau) = e^{i\hat{H}_{0,\mu}\tau} \hat{\mu} e^{-i\hat{H}_{0,\mu}\tau}$$

with $\hat{H}_{0,\mu}|E_i\rangle = E_i |E_i\rangle$ we obtain for the transition amplitude at time t_f

$$i\lambda \langle E|\mu(0)|E_0\rangle \int_{\tau_i}^{\tau_f} \mathrm{d}\tau \mathrm{e}^{\mathrm{i}(E-E_0)\tau} \langle\psi|\hat{\phi}(x(\tau))|0\rangle \,. \tag{B.27}$$

In our limited treatment, we do not model the dynamics of the black hole and control variable or apparatus keeping the detector on its trajectory. As a result, we do not acquire a branch-dependent phase associated with these degrees of freedom. If we have a single path or geometry, an overall phase Φ is not important, but when we have a superposition of amplitudes a relative phase $\Delta\Phi$ might contribute. The relative phase $\Delta\Phi$ in our case may have several potential origins or contributions in a more complete treatment. First, it may arise from the approximate evolution of the black hole as an approximate eigenstate of the gravitational Hamiltonian with its mass as eigenvalue with respect to coordinate time t if we were to include the corresponding Hilbert space and state in our initial-state description; the gravitational Hamiltonian would then become a part of $\hat{H}^{\tau}_{0,S}$ in Eqs. (B.23) to (B.25), taking into account that $\hat{H}^{\tau}_{\text{grav}} = \frac{dt}{d\tau}\hat{H}^{t}_{\text{grav}}$. A relative phase difference when considering proper-time evolution might then appear as coordinate time relates differently to the detector's proper time in the two branches. The precise assumption in that case would thus be that for the purpose of the experiment the BTZ black hole spacetime remains approximately static and is an approximate eigenstate of the gravitational Hamiltonian, so that the phase coming from the black hole evolution is determined by its eigenvalue, presumably proportional to the black hole's ADM mass M (in line with [119]) i.e. $H^t_0 |M\rangle_{BH} = M |M\rangle_{BH}$. Such phase might thus also be particularly relevant in the case of a mass-superposed black hole where a relative phase

between the two branches of the superposition can also develop as the mass differs between the two branches. As we consider our black hole to have the same mass in both branches of the superposition, we do not assign an explicit Hilbert space for the mass to the black hole, only for its position. In a more complete treatment, the spacetime location and energy-momentum of the black hole may not be independent [452, 609]. Approximating the black hole as developing only a phase independent of spatial location might also be an issue because of noncovariance, which would require a dynamical analysis taking into account the gravitational constraints and backreaction, potential non-factorization, etc. We neglect these complications here. Secondly, we also neglect any additional relative phase associated with the external control degree of freedom and the apparatus responsible for keeping the detector on the prescribed static trajectory. If one were to take into account these phases, we could also simply absorb them in the definition of the states. Whether an additional phase would be visible in the experimental data depends on the precise implementation of the interferometric black hole position measurement which we describe below.

Next we consider a detector in a superposition of classical geometries with black hole at positions $\mathbf{x}_1, \mathbf{x}_2$ with total initial state in the detector frame given by $|\psi(\tau_i)\rangle_S = |\mathbf{0}\rangle_D (|\mathbf{x}_1\rangle_B + |\mathbf{x}_2\rangle_B) |0\rangle_\phi |E_0\rangle_\mu/\sqrt{2}$ with $|0\rangle_\phi$ denoting the scalar field vacuum. In the case of a superposition of geometries (or detector location under a QRF transformation) one needs to be careful regarding coordinates, especially the choice of time parametrization used. In the detector QRF the proper time τ is single-valued and well-defined, but the black hole is in a superposition of locations so that the relation between proper time and coordinate times differs, whereas in the QRF with the black hole at the origin the coordinate time t is well-defined but the proper time of the detector is indefinite. The relation between coordinate time t and proper time τ is given by the time dilation of Eq. (9.2), $\tau = t\gamma(R)$ with R the radial coordinate distance between the detector and black hole.

In the detector frame (i.e. putting the detector at the origin), the black hole is in a superposition of locations. Taking the interaction picture representation of the QRF-transformed UDW interaction Eq. (9.24) and post-selecting on the detector making a transition to energy level E we obtain at time τ_f

$$\frac{1}{\sqrt{2}}\Big(|\mathbf{x}_1\rangle_B \Theta_1 + |\mathbf{x}_2\rangle_B \Theta_2\Big), \tag{B.28}$$

where

$$\Theta_J = i\lambda\langle E|\hat{\mu}|E_0\rangle \int_{\tau_i}^{\tau_f} \mathrm{d}\tau\eta(\tau)e^{i(E-E_0)\tau}\langle\vartheta|\hat{\phi}_J(x_J(\tau))|0\rangle \tag{B.29}$$

with ϕ_J the scalar field coupled to the metric g_J with black hole at position $\mathbf{x}_J$ and $|\vartheta\rangle$ the final state of the scalar field which we will sum over. Note that if another clock were used to turn the detectors on and off, the proper time intervals for the UDW interaction could differ between the two branches. We have omitted the detector's position state here as we are in the frame where this position is factorised as being the origin. As explained below Eq. (B.27), no additional branch-dependent relative phase is generated. Given the time-localized switching function, this also means that we could as well take $\tau_f, \tau_i \to \pm\infty$. If we now consider a joint measurement of the detector transitioning to the excited state $|E\rangle_\mu$ at time τ_f and an interferometric measurement on the black hole position, i.e. measuring in the basis $(|\mathbf{x}_1\rangle \pm |\mathbf{x}_2\rangle)_B/\sqrt{2}$, we obtain the post-selected detector response given in Eq. (9.37).

B.4 Calculating integrals for the measurement probabilities

B.4.1 Calculating the cross-term $\mathcal{F}_{12}(\Omega)$

We now insert the BTZ correlation function Eq. (B.10) (for transparent boundary conditions, calculations for other boundary conditions are analogous) into the cross-term for the probability Eq. (9.38), defining a response function $\mathcal{F}_{12} = P_{12}/\mathcal{N}'$ with $\mathcal{N}' = \lambda^2|\langle E|\mu|E_0\rangle|^2$, obtaining[1]

$$\mathcal{F}_{12}(\Omega) = \int_{\tau_i}^{\tau_f} d\tau \int_{\tau_i}^{\tau_f} d\tau'\eta(\tau)\eta(\tau')e^{-i\Omega(\tau-\tau')}W_{BTZ}(x_1(\tau), x_2(\tau')) \tag{B.30}$$

where the paths $x_1(\tau) = (\tau/\gamma_1, R_1, 0), x_2(\tau') = (\tau'/\gamma_2, R_2, \theta)$ with $\gamma_i = \gamma(R_i)$ and $\Omega = \Delta E = E - E_0$. Now we move to coordinate time, obtaining

$$\begin{aligned}\mathcal{F}_{12}(\Omega) &= \gamma_1\gamma_2 \int_{\tau_i/\gamma_1}^{\tau_f/\gamma_1} dt \int_{\tau_i/\gamma_2}^{\tau_f/\gamma_2} dt'\eta(\gamma_1 t)\eta(\gamma_2 t')e^{-i\Omega(\gamma_1 t-\gamma_2 t')}W_{BTZ}(x_1(\tau), x_2(\tau')) \\ &= \gamma_1\gamma_2 \int_{-\tau_f/\gamma_1}^{\tau_f/\gamma_1} du \int_{-\tau_f/\gamma_2+u}^{\tau_f/\gamma_2+u} ds\eta(\gamma_1 u)\eta(\gamma_2(u-s))e^{-i\Omega\left(\gamma_1 u-\gamma_2(u-s)\right)}W^{12}_{BTZ}(s)\end{aligned} \tag{B.31}$$

[1]The calculations here resemble those of the Supplemental Material of Ref. [119] closely, but are slightly different as we consider a position-superposed BTZ black hole, as opposed to the mass-superposed case of Ref. [119].

where we have chosen $\tau_i = -\tau_f$ for simplicity, and changed variables $u = t$, $s = t - t'$ and $W^{12}_{BTZ}(t-t')$ is defined in Eq. (B.10). Next, we make the same approximation as in Appendix D (S31) of [119]:

$$\begin{aligned}\int_{-t_{f1}}^{t_{f1}} \mathrm{d}u \int_{u-t_{f2}}^{u+t_{f2}} \mathrm{d}s\eta\left(\gamma_D u\right)\eta\left(\gamma_D(u-s)\right)\ldots \\ \simeq \int_{-t_{f1}}^{t_{f1}} \mathrm{d}u \int_{-t_{f2}}^{t_{f2}} \mathrm{d}s\eta\left(\gamma_D u\right)\eta\left(\gamma_D(u-s)\right)\ldots,\end{aligned} \tag{B.32}$$

with $t_{fJ} = \tau_f/\gamma_J$. This approximation would be exact if the integration limits were $-\infty$ to ∞; recall that we required that $\Delta\tau \gg \sigma$ such that the switching functions have small support outside $[-\tau_f/\gamma_J, \tau_f/\gamma_J]$. Using Eq. (B.32) in Eq. (B.31), we find

$$\begin{aligned}\mathcal{F}_{12}(\Omega) &\approx \gamma_1\gamma_2 \int_{-\tau_f/\gamma_1}^{\tau_f/\gamma_1} du \int_{-\tau_f/\gamma_2}^{\tau_f/\gamma_2} ds\eta(\gamma_1 u)\eta(\gamma_2(u-s))e^{-i\Omega\left(\gamma_1 u-\gamma_2(u-s)\right)}W^{12}_{BTZ}(s) \\ &= \gamma_1\gamma_2 \int_{-\tau_f/\gamma_2}^{\tau_f/\gamma_2} ds I(s) e^{-\frac{\gamma_2^2 s^2}{2\sigma^2}} e^{-i\Omega\gamma_2 s} W^{12}_{BTZ}(s)\end{aligned} \tag{B.33}$$

where $I(s)$ is an integral over u:

$$\begin{aligned}I(s) &= \int_{-\tau_f/\gamma_1}^{\tau_f/\gamma_1} du e^{-\frac{\gamma_1^2 u^2}{2\sigma^2}} e^{-\frac{\gamma_2^2 u^2 - 2\gamma_2^2 us}{2\sigma^2}} e^{-i\Omega(\gamma_1-\gamma_2)u} \\ &= \int_{-\tau_f/\gamma_1}^{\tau_f/\gamma_1} du e^{-\frac{\left(\sqrt{\gamma_1^2+\gamma_2^2}u - \frac{\gamma_2^2}{\sqrt{\gamma_1^2+\gamma_2^2}}s + i\Omega\frac{(\gamma_1-\gamma_2)}{\sqrt{\gamma_1^2+\gamma_2^2}}\sigma^2\right)^2}{2\sigma^2}} e^{\frac{\gamma_2^4 s^2}{2\sigma^2(\gamma_1^2+\gamma_2^2)}} e^{-\Omega^2\sigma^2\frac{(\gamma_1-\gamma_2)^2}{2(\gamma_1^2+\gamma_2^2)}} e^{i\Omega\frac{\gamma_2(\gamma_1-\gamma_2)s}{\gamma_1^2+\gamma_2^2}} \\ &= \frac{\sqrt{2}\sigma}{\sqrt{\gamma_1^2+\gamma_2^2}} e^{\frac{\gamma_2^4 s^2}{2\sigma^2(\gamma_1^2+\gamma_2^2)}} e^{-\Omega^2\sigma^2\frac{(\gamma_1-\gamma_2)^2}{2(\gamma_1^2+\gamma_2^2)}} e^{-i\Omega\frac{\gamma_2^2(\gamma_1-\gamma_2)s}{\gamma_1^2+\gamma_2^2}} \\ &\int_{-\frac{\sqrt{\gamma_1^2+\gamma_2^2}\tau_f}{\sqrt{2}\sigma\gamma_1} - \frac{\gamma_2^2 s}{\sqrt{\gamma_1^2+\gamma_2^2}\sqrt{2}\sigma} + i\Omega\sigma\frac{\gamma_1-\gamma_2}{\sqrt{2}\sqrt{\gamma_1^2+\gamma_2^2}}}^{\frac{\sqrt{\gamma_1^2+\gamma_2^2}\tau_f}{\sqrt{2}\sigma\gamma_1} - \frac{\gamma_2^2 s}{\sqrt{\gamma_1^2+\gamma_2^2}\sqrt{2}\sigma} + i\Omega\sigma\frac{\gamma_1-\gamma_2}{\sqrt{2}\sqrt{\gamma_1^2+\gamma_2^2}}} dz e^{-z^2} \\ &=: \frac{\sqrt{2}\sigma}{\sqrt{\gamma_1^2+\gamma_2^2}} e^{\frac{\gamma_2^4 s^2}{2\sigma^2(\gamma_1^2+\gamma_2^2)}} e^{-\Omega^2\sigma^2\frac{(\gamma_1-\gamma_2)^2}{2(\gamma_1^2+\gamma_2^2)}} e^{-i\Omega\frac{\gamma_2^2(\gamma_1-\gamma_2)s}{\gamma_1^2+\gamma_2^2}} \frac{\sqrt{\pi}}{2} H(s).\end{aligned} \tag{B.34}$$

Here we changed variables $z = u\frac{\sqrt{\gamma_1^2+\gamma_2^2}}{\sqrt{2}\sigma} + \ldots$ and $H(s)$ is given by

$$\begin{aligned}H(s) =& \mathrm{erf}\left(\frac{(\gamma_1^2+\gamma_2^2)\tau_f - \gamma_1\gamma_2^2 s + i\Omega\sigma^2\gamma_1(\gamma_1-\gamma_2)}{\sqrt{2}\sigma\gamma_1\sqrt{\gamma_1^2+\gamma_2^2}}\right) \\ &+ \mathrm{erf}\left(\frac{(\gamma_1^2+\gamma_2^2)\tau_f + \gamma_1\gamma_2^2 s - i\Omega\sigma^2\gamma_1(\gamma_1-\gamma_2)}{\sqrt{2}\sigma\gamma_1\sqrt{\gamma_1^2+\gamma_2^2}}\right)\end{aligned} \tag{B.35}$$

with erf the Gaussian error function. Using Eq. (B.34) in Eq. (B.33) we find

$$\begin{aligned}\mathcal{F}_{12}(\Omega) &\approx \frac{1}{4\pi l\sqrt{2}}\frac{\sqrt{M}}{\sqrt{\gamma_1\gamma_2}}\frac{\gamma_1\gamma_2\sigma}{\sqrt{\gamma_1^2+\gamma_2^2}}\sqrt{\frac{\pi}{2}}e^{-\Omega^2\sigma^2\frac{(\gamma_1-\gamma_2)^2}{2(\gamma_1^2+\gamma_2^2)}} \\ &\qquad \sum_n \int_{-\tau_f/\gamma_2}^{\tau_f/\gamma_2} ds \frac{e^{-\frac{\gamma_1^2\gamma_2^2}{2\sigma^2(\gamma_1^2+\gamma_2^2)}s^2} e^{-iEs\frac{\gamma_1\gamma_2(\gamma_1+\gamma_2)}{\gamma_1^2+\gamma_2^2}} H(s)}{\sqrt{\alpha_n^{(12)} - \cosh\left(\frac{r_h}{l^2}s - i\epsilon\right)}} \\ &= \sigma\frac{\sqrt{M}}{2l}Y_0 \sum_n \int_{-\tau_f/\gamma_2}^{\tau_f/\gamma_2} ds \frac{Z_0(s)H(s)}{\sqrt{\alpha_n^{(12)} - \cosh\left(\frac{r_h}{l^2}s - i\epsilon\right)}} \\ &= \sigma Y_0 \sum_n \text{ Re} \int_0^{\frac{\tau_f\sqrt{M}}{l\gamma_2}} dz \frac{Z_0(lz/\sqrt{M})H(lz/\sqrt{M})}{\sqrt{\alpha_n^{(12)} - \cosh(z - i\epsilon)}}\end{aligned} \tag{B.36}$$

with $H(s)$ given in Eq. (B.35), α_n^{12} given in Eq. (B.11) and Y_0 and $Z_0(s)$ given by

$$Y_0 = \frac{\sqrt{\gamma_1\gamma_2}\sqrt{\pi}}{4\pi\sqrt{\gamma_1^2+\gamma_2^2}}e^{-\Omega^2\sigma^2\frac{(\gamma_1-\gamma_2)^2}{2(\gamma_1^2+\gamma_2^2)}}, \tag{B.37}$$

$$Z_0(s) = e^{-\frac{\gamma_1^2\gamma_2^2}{2\sigma^2(\gamma_1^2+\gamma_2^2)}s^2} e^{-i\Omega s\frac{\gamma_1\gamma_2(\gamma_1+\gamma_2)}{\gamma_1^2+\gamma_2^2}}. \tag{B.38}$$

The cross-term will typically also contain singularities, but integrable ones.

For Neumann and Dirichlet boundary conditions, we have an extra term, namely

$$\begin{aligned}\mathcal{F}_{12}(\Omega) = \sigma Y_0 \sum_n \text{ Re} \int_0^{\frac{\tau_f\sqrt{M}}{l\gamma_2}} dz Z_0(lz/\sqrt{M})H(lz/\sqrt{M}) \Bigg(&\frac{1}{\sqrt{\alpha_n^{(12)} - \cosh(z - i\epsilon)}} \\ &\pm \frac{1}{\sqrt{\alpha_n'^{12} - \cosh(z - i\epsilon)}}\Bigg)\end{aligned} \tag{B.39}$$

with $+, -$ for Neumann and Dirichlet boundary conditions, respectively.

B.4.2 Calculating $\mathcal{F}_J(E)$

The integrals for the diagonal terms $\mathcal{F}_J(E)$ are exactly the ones of [119], we shortly repeat those here as slightly different normalizations were used; they are analogous to calculation of $\mathcal{F}_{12}(\Omega)$ just presented. We omit the label $J = 1, 2$ here as we are now in the case of a single non-superposed black hole and localized detector.

The response function is given by

$$\mathcal{F}(\Omega) = \int_{\tau_i}^{\tau_f} d\tau \int_{\tau_i}^{\tau_f} d\tau' \eta(\tau)\eta(\tau') e^{-i\Omega(\tau-\tau')} W_{BTZ}(x(\tau), x(\tau')). \tag{B.40}$$

We first move to coordinate time, obtaining

$$\begin{aligned}\mathcal{F}(\Omega) =& \gamma^2 \int_{\tau_i/\gamma}^{\tau_f/\gamma} dt \int_{\tau_i/\gamma}^{\tau_f/\gamma} dt' \eta(\gamma t)\eta(\gamma t') e^{-i\Omega\gamma(t-t')} W_{BTZ}(x(\gamma t), x(\gamma t')) \\ =& \gamma^2 \int_{\tau_i/\gamma}^{\tau_f/\gamma} du \int_{\tau_i/\gamma+u}^{\tau_f/\gamma+u} ds \eta(\gamma u)\eta(\gamma(u-s)) e^{-i\Omega\gamma s} W_{BTZ}(s)\end{aligned} \tag{B.41}$$

with $\gamma = \gamma(R) = \sqrt{-g_{00}}$, we chose $\tau_i = -\tau_f$, and we have performed the change of variables $u = t, s = t - t'$. Next, we require $\Delta\tau >> \sigma$ such that we can use the approximation of Equation (B.32), taking $\tau_i = -\tau_f$ giving

$$\begin{aligned}\mathcal{F}(\Omega) \approx& \gamma^2 \int_{-\tau_f/\gamma}^{\tau_f/\gamma} du \int_{-\tau_f/\gamma}^{\tau_f/\gamma} ds \eta(\gamma u)\eta(\gamma(u-s)) e^{-i\Omega\gamma s} W_{BTZ}(s) \\ =& \gamma^2 \int_{-\tau_f/\gamma}^{\tau_f/\gamma} ds J(s) e^{-\gamma^2 \frac{s^2}{4\sigma^2}} e^{-i\Omega\gamma s} W_{BTZ}(s)\end{aligned} \tag{B.42}$$

with

$$\begin{aligned}J(s) &= \int_{-\tau_f/\gamma}^{\tau_f/\gamma} du e^{-\gamma^2 \frac{(u-1/4s)^2}{\sigma^2}} \\ &= \frac{\sigma\sqrt{\pi}}{2\gamma}\left(\operatorname{erf}\frac{\tau_f - \gamma s/2}{\sigma} + \operatorname{erf}\frac{\tau_f + \gamma s/2}{\sigma}\right) \\ &= \frac{\sigma\sqrt{\pi}}{2\gamma} H_0(s).\end{aligned} \tag{B.43}$$

Here we defined

$$H_0(s) = \operatorname{erf}\frac{\tau_f - \gamma s/2}{\sigma} + \operatorname{erf}\frac{\tau_f + \gamma s/2}{\sigma}. \tag{B.44}$$

If we take the integration limits τ_f to infinity we can approximate this integral $J(s)$ simply by $\sigma\sqrt{\pi}/\gamma$.

We thus find that

$$\begin{aligned}\mathcal{F}(\Omega) &\approx \frac{\gamma\sigma\sqrt{\pi}}{2} \int_{-\tau_f/\gamma}^{\tau_f/\gamma} ds H_0(s) e^{-\gamma^2 \frac{s^2}{4\sigma^2}} e^{-i\Omega\gamma s} W_{BTZ}(s) \\ &= \frac{\gamma\sigma\sqrt{\pi}}{2} \frac{\sqrt{M}}{4\pi\gamma l\sqrt{2}} \sum_m \int_{-\tau_f/\gamma}^{\tau_f/\gamma} ds \frac{H_0(s) e^{-\gamma^2 \frac{s^2}{4\sigma^2}} e^{-i\Omega\gamma s}}{\sqrt{\alpha_m - \cosh\left(\frac{r_h}{l^2}s - i\epsilon\right)}} \\ &= \frac{\sigma\sqrt{M}}{8l\sqrt{2\pi}} \sum_m \int_{-\tau_f/\gamma}^{\tau_f/\gamma} ds \frac{H_0(s) e^{-\gamma^2 \frac{s^2}{4\sigma^2}} e^{-i\Omega\gamma s}}{\sqrt{\alpha_m - \cosh\left(\frac{r_h}{l^2}s - i\epsilon\right)}}\end{aligned} \tag{B.45}$$

where we have inserted the BTZ correlation function of Equation (B.10) and α_m is given by

$$\alpha_m = \frac{R^2}{\gamma^2 l^2}\cosh\left(2\pi m\sqrt{M}\right) - \frac{M}{\gamma^2}. \tag{B.46}$$

Note that $Y_0, Z_0(lz)$ are the same as for the cross term for the superposed black hole mass case of Ref. [119], but $H(lz)$ and α_n^{12} are different.

B.4.3 Relating to the coordinates of Ref. [119] and final expressions

The calculations in Ref. [119] were done in a rescaled AdS time coordinate. Namely, Ref. [119] works with the time coordinates $\bar{t}$ of AdS_3, of which the BTZ black hole is obtained by a periodic identification as explained in Section B.1. This means that our coordinates t, R are related to the coordinate $\bar{t}, \overline{R}$ of 3d AdS–Rindler space of Section B.1 by

$$\begin{aligned} \sqrt{M}t &= \bar{t}, \\ R/\sqrt{M} &= \overline{R}. \end{aligned} \tag{B.47}$$

Our coordinates t, R are referred to as $\tilde{s}, \tilde{R}$ in the Supplemental Materials of Ref. [119], but the tildes are omitted there in the Supplemental Sections D and E. Ref. [119] also defines a rescaled redshift factor which we refer to as $\tilde{\gamma}$, defined by

$$\tilde{\gamma}(R) = \sqrt{\frac{R^2}{Ml^2} - 1} = \sqrt{\frac{\bar{R}^2}{l^2} - 1}, \tag{B.48}$$

and thus related to our BTZ redshift factor γ of Eq. (9.2) by

$$\gamma(R) = \tilde{\gamma}(R)\sqrt{M}. \tag{B.49}$$

Note that Ref. [119] still uses the BTZ radial coordinate. Furthermore, because Ref. [119] rescaled the times by $\sqrt{M}$, the time span in $\tilde{t_f}$ is equal to the time span in coordinate t times $\sqrt{M}$, so that, if one were to take a gravitational phase into account, our definition of the phase being $M \cdot t$ is dimensionally consistent with their definition of the phase being $\sqrt{M} \cdot \bar{t}$.

We can put our results for $\mathcal{F}_{12}(\Omega)$ and $\mathcal{F}(\Omega)$ in terms of the coordinates and conventions of Ref. [119] as follows. In the remaining part of this section, t, γ should be understood as $\bar{t}, \tilde{\gamma}$, finding

$$\frac{\mathcal{F}_{12}(\Omega)}{\sigma} \approx Y_0 \sum_n \ \mathrm{Re}\int_0^{\frac{\tau_f}{l\tilde{\gamma}_2}} dz \frac{Z_0(lz)H(lz)}{\sqrt{\alpha_n - \cosh(z - i\epsilon)}} \tag{B.50}$$

with

$$\alpha_n^{12} = \frac{R_1 R_2}{\widetilde{\gamma}_1\widetilde{\gamma}_2\sqrt{M_1 M_2 l^2}}\cosh\left(2\pi n\sqrt{M}\right) - \frac{1}{\widetilde{\gamma}_1\widetilde{\gamma}_2} \tag{B.51}$$

$$Y_0 = \frac{\sqrt{\widetilde{\gamma}_1\widetilde{\gamma}_2}\sqrt{\pi}}{4\pi\sqrt{\widetilde{\gamma}_1^2+\widetilde{\gamma}_2^2}} e^{-\Omega^2\sigma^2\frac{(\widetilde{\gamma}_1-\widetilde{\gamma}_2)^2}{2(\widetilde{\gamma}_1^2+\widetilde{\gamma}_2^2)}}, \tag{B.52}$$

$$Z_0(s) = e^{-\frac{\widetilde{\gamma}_1^2\widetilde{\gamma}_2^2}{2\sigma^2(\widetilde{\gamma}_1^2+\widetilde{\gamma}_2^2)}s^2} e^{-i\Omega s\frac{\widetilde{\gamma}_1\widetilde{\gamma}_2(\widetilde{\gamma}_1+\widetilde{\gamma}_2)}{\widetilde{\gamma}_1^2+\widetilde{\gamma}_2^2}}, \tag{B.53}$$

$$H(s) = \mathrm{erf}\left(\frac{(\widetilde{\gamma}_1^2+\widetilde{\gamma}_2^2)\tau_f - \widetilde{\gamma}_1\widetilde{\gamma}_2^2 s + i\Omega\sigma^2\widetilde{\gamma}_1(\widetilde{\gamma}_1-\widetilde{\gamma}_2)}{\sqrt{2}\sigma\widetilde{\gamma}_1\sqrt{\widetilde{\gamma}_1^2+\widetilde{\gamma}_2^2}}\right) \tag{B.54}$$

$$+\,\mathrm{erf}\left(\frac{(\widetilde{\gamma}_1^2+\widetilde{\gamma}_2^2)\tau_f + \widetilde{\gamma}_1\widetilde{\gamma}_2^2 s - i\Omega\sigma^2\widetilde{\gamma}_1(\widetilde{\gamma}_1-\widetilde{\gamma}_2)}{\sqrt{2}\sigma\widetilde{\gamma}_1\sqrt{\widetilde{\gamma}_1^2+\widetilde{\gamma}_2^2}}\right) \tag{B.55}$$

and $\gamma_i(R) = \sqrt{R_i^2/(Ml^2) - 1}$ for $i = 1, 2$. If $\operatorname{arccosh}\alpha_n < \tau_f/(l\widetilde{\gamma}_2)$, we need to make sure to take the correct branch of the square root, i.e.

$$\begin{aligned}\frac{\mathcal{F}_{12}(\Omega)}{\sigma} \approx & Y_0\sum_n \mathrm{Re}\int_0^{\operatorname{arccosh}(\alpha_n)} dz \frac{Z_0(lz)H(lz)}{\sqrt{\alpha_n - \cosh(z-i\epsilon)}} \\ & + Y_0\sum_n \mathrm{Re}\int_{\operatorname{arccosh}(\alpha_n)}^{\tau_f/(l\widetilde{\gamma}_2)} dz \frac{Z_0(lz)H(lz)}{i\sqrt{\cosh(z-i\epsilon) - \alpha_n}}\end{aligned} \tag{B.56}$$

For $\mathcal{F}(\Omega)$ from Equation (B.45), we take a single branch with detector at radial coordinate distance R from the black hole (so $R = R_1, R_2$ for the two branches of our position-superposition), and $\widetilde{\gamma}$ is the related (rescaled) redshift, so $\widetilde{\gamma} = \widetilde{\gamma}_1, \widetilde{\gamma}_2$ for the two branches. Using the rescaled coordinates of Ref. [119] as explained above, we obtain

$$\mathcal{F}(\Omega) \approx \frac{\sigma}{8l\sqrt{2\pi}}\sum_m \int_{-\tau_f/\widetilde{\gamma}}^{\tau_f/\widetilde{\gamma}} ds \frac{H_0(s) e^{-\widetilde{\gamma}^2\frac{s^2}{4\sigma^2}} e^{-i\Omega\widetilde{\gamma}s}}{\sqrt{\alpha_m - \cosh(s/l - i\epsilon)}}, \tag{B.57}$$

which, as shown in Ref. [119], leads to the following

$$\begin{aligned}\frac{\mathcal{F}(\Omega)}{\sigma} \approx & \frac{\sqrt{\pi}H_0(0)}{8} - \frac{i}{8\sqrt{\pi}}\mathrm{PV}\int_{-\tau_f/(2l\widetilde{\gamma})}^{\tau_f/(2l\widetilde{\gamma})} \mathrm{d}z \frac{X_0(2lz)H_0(2lz)}{\sinh(z)} \\ & + \frac{1}{4\sqrt{2\pi}}\sum_{n\neq 0}\mathrm{Re}\int_0^{\tau_f/(l\widetilde{\gamma})} \mathrm{d}z \frac{X_0(lz)H_0(lz)}{\sqrt{\alpha_n - \cosh(z)}}.\end{aligned} \tag{B.58}$$

with

$$X_0(s) = e^{-\frac{\widetilde{\gamma}^2 s^2}{4\sigma^2}} e^{-i\Omega\widetilde{\gamma}s}, \tag{B.59}$$

$$\alpha_n = \frac{R^2}{M\widetilde{\gamma}^2 l^2}\cosh\left(2\pi n\sqrt{M}\right) - \frac{1}{\widetilde{\gamma}^2}, \tag{B.60}$$

$$H_0(s) = \operatorname{erf}\left[\frac{\widetilde{\gamma}\,(s + 2\tau_f/\widetilde{\gamma})}{2\sigma}\right] - \operatorname{erf}\left[\frac{\widetilde{\gamma}\,(s - 2\tau_f/\widetilde{\gamma})}{2\sigma}\right]. \tag{B.61}$$

The first two terms of Equation (B.58) arise from a pole at $n = 0$ using the Sokhotski formula

$$\frac{1}{\sinh(z - i\epsilon)} = i\pi\delta(z) + \mathrm{PV}\frac{1}{\sinh(z)}. \tag{B.62}$$

Here we have corrected a small inconsistency in the Supplemental Material of Ref. [119]: in Eq. (S51) of Ref. [119], the first term in β_{nm} should be divided by M, i.e. the expression of β_{nm} should be

$$\beta_{nm} = \frac{R_D^2}{M\widetilde{\gamma}_D^2 l^2}\cosh\left(2\pi(n-m)\sqrt{M}\right) - \frac{1}{\widetilde{\gamma}_D^2} \tag{B.63}$$

where we have highlighted the missing term M in blue. The expression β_{n0} corresponds to our α_n in Eq. (B.60).

Similarly to Eq. (B.56), if the term inside the square root in the denominator of the third term in Eq. (B.57) becomes negative, taking care of the square root leads to

$$\frac{1}{4\sqrt{2\pi}}\left[\sum_{n\neq 0}\mathrm{Re}\int_0^{\operatorname{arccosh}\alpha_n}\mathrm{d}z\frac{X_0(lz)H_0(lz)}{\sqrt{\alpha_n - \cosh(z)}} + \sum_{n\neq 0}\mathrm{Re}\int_{\operatorname{arccosh}\alpha_n}^{\tau_f/(l\widetilde{\gamma})}\mathrm{d}z\frac{X_0(lz)H_0(lz)}{i\sqrt{\cosh(z) - \alpha_n}}\right]. \tag{B.64}$$

B.5 Derivation of interaction Hamiltonian for superposed black hole

In this section we derive the form of the Unruh–DeWitt detector Hamiltonian in the laboratory frame, i.e. in the presence of a superposition of black holes. The quantum-controlled UDW interaction introduced in Section 9.2.2 in the black hole frame (but according to the detector's time) is, in the Schrödinger picture, given by

$$H_{\mathsf{int}}^{(B)} = \lambda\,\gamma(\hat{\mathbf{x}})\,\mu\otimes\phi(\hat{\mathbf{x}}) = \lambda\,\mu\otimes\int d\nu(\mathbf{x})\;|\mathbf{x}\rangle\langle\mathbf{x}|_D\;\gamma(\mathbf{x})\otimes\phi(\mathbf{x}) \tag{B.65}$$

acting on the Hilbert space $\mathcal{H}_D \otimes \mathcal{H}_\mu \otimes \mathcal{H}_\phi$ corresponding to the external degrees of freedom of the detector (its position), the internal degrees of freedom of the detector (the n-level energy system) and the scalar field sector, respectively. Recall that $\gamma(\mathbf{x})$ is the redshift factor. A QRF transformation maps the perspective of the black hole to the perspective of the detector (i.e. the laboratory), with the transformation

$$S^{(B\to D)} = \int d\nu(\mathbf{x}) \; |-\mathbf{x}\rangle_B \langle \mathbf{x}|_D \otimes \mathbb{I}_\phi \otimes \mathbb{I}_\mu \tag{B.66}$$

After the QRF transformation, the UDW interaction Hamiltonian in the detector frame becomes

$$\begin{aligned} H_{\text{int}}^{(D)} &= S^{(B\to D)} \gamma^{-1}(\hat{\mathbf{x}}_D) H_{\text{int}}^{(B)} S^{\dagger(B\to D)} = \lambda \int d\nu(\mathbf{x}) \; |\mathbf{x}\rangle \langle \mathbf{x}|_B \otimes \mu \otimes \phi(-\mathbf{x}) \\ &= \lambda \int d\nu(\mathbf{x}) \; |\mathbf{x}\rangle \langle \mathbf{x}|_B \otimes H_{int,\mathbf{x}_B}, \end{aligned} \tag{B.67}$$

with $H_{int,\mathbf{x}_B}$ the Unruh–DeWitt interaction for a detector with a scalar field in a background with black hole at position $\mathbf{x}$ (and with detector now at the origin $\mathbf{0}$).

B.6 Derivation of the analytical expressions of Section 9.4.3

B.6.1 Summary of expressions

In this section we present the analytical expressions for the spectrum probed by the detector from Section 9.4.3. Before presenting detailed derivations, we provide a summary of the relevant expressions here. We consider a static detector near a (potentially superposed) BTZ black hole.

Single-spacetime detector response. In a non-superposed spacetime, the detector response from Eq. (9.11) can be expressed as

$$\mathcal{F}(E) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega |\widehat{\eta}(\tfrac{\omega}{\gamma})|^2 \widehat{W}_{BTZ}(E\gamma + \omega), \tag{B.68}$$

The explicit form of the BTZ correlation function along the detector's path is given by taking $R_1 = R_2 = R$ to be the detector's radial coordinate and $\theta = 0$ in Eq. (B.12):

$$W_{BTZ}(s) = \frac{1}{4\pi l\sqrt{2\widetilde{\gamma}}} \left[\frac{1}{\sqrt{\alpha_m - \cosh\left(s\sqrt{M}/l - i\epsilon\right)}} - \frac{\zeta}{\sqrt{\alpha'_m - \cosh\left(s\sqrt{M}/l - i\epsilon\right)}} \right] \tag{B.69}$$

with limit $\epsilon \to 0^+$, $s = t - t', \tilde{\gamma} = \gamma(R)/\sqrt{M}$ and

$$\alpha_m = \frac{\frac{R^2}{Ml^2}\cosh[2\pi m\sqrt{M}] - 1}{\frac{R^2}{Ml^2} - 1} \tag{B.70}$$

$$\alpha'_m = \frac{\frac{R^2}{Ml^2}\cosh[2\pi m\sqrt{M}] + 1}{\frac{R^2}{Ml^2} - 1}. \tag{B.71}$$

Following Ref. [746], an analytical expression for $\widehat{W}$ can be found, giving[2]

$$\widehat{W}_{BTZ}(K) = \frac{1}{4\gamma} + \frac{1}{2\gamma}\frac{1}{e^{K/T_H}+1}\sum_n \left[P_{\frac{iK}{2\pi T_H}-\frac{1}{2}}(\alpha_n) - \zeta P_{\frac{iK}{2\pi T_H}-\frac{1}{2}}(\alpha'_n)\right] \tag{B.72}$$

with Hawking temperature $T_H = \sqrt{M}/(2\pi l)$ and P_ν the Legendre function of the first kind.

Interference term in detector response. The interference term $\mathcal{F}_{12}(E)$, for both mass- and position-superposed black holes now, can be written in a similar form as Eq. (B.68):

$$\mathcal{F}_{12}(E) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{d}\omega\, \widehat{\eta}(\tfrac{\omega}{\gamma_1})^* \,\widehat{\eta}(\tfrac{\omega}{\gamma_2} + E[\tfrac{\widetilde{\gamma_1}}{\gamma_2} - 1])\, \widehat{W}^{(12)}_{BTZ}(\omega + \widetilde{\gamma_1}E). \tag{B.73}$$

To calculate the Fourier transform of $\widehat{W}^{(12)}$, we will first perform a change of coordinates in each branch of the superposition. Namely, we want to be able to treat also the mass-superposed case, in which case we need coordinates to compare spacetimes for different black hole masses. Following Ref. [119], we do so by considering AdS time coordinates $\bar{t}$ related to the BTZ time coordinates t by $\bar{t} = t\sqrt{M_i}$ in branch i. In the mass-superposed black hole, we now need the additional assumption that the Unruh–DeWitt interaction is a quantum-controlled version of the single-spacetime interaction, with control on the black hole properties, its position and mass, as in Ref. [119]. We have made this assumption more plausible by deriving this interaction form in the case of the black hole position in Eq. (9.24). We denote the position and mass by R_1, M_1 and R_2, M_2 in the two branches of the superposition. We also introduce $\widetilde{\gamma_i}(R_i) = \sqrt{R_i^2/(M_i l^2) - 1}$.

The scalar field correlation function between two BTZ spacetimes is then given by [119]:

$$\widehat{W}^{(12)}_{BTZ}(\bar{s}) = \frac{1}{4\pi\sqrt{2\widetilde{\gamma_1}\widetilde{\gamma_2}}}\frac{1}{\sum_k \Upsilon^{2k}}\sum_{m,n}\left[\frac{1}{\sqrt{\beta^{(12)}_{mn} - \cosh(\bar{s}/l)}} - \frac{\zeta}{\sqrt{\beta'^{(12)}_{mn} - \cosh(\bar{s}/l)}}\right], \tag{B.74}$$

[2] Compared to Ref. [746] we have added the first term arising from the pole singularity using the Sokhotski formula.

with $\bar{s} = \bar{t} - \bar{t}' - i\epsilon$ and

$$\beta_{mn}^{(12)} = \frac{\frac{R_1 R_2}{\sqrt{M_1 M_2} l^2} \cosh[2\pi(m\sqrt{M_1} - n\sqrt{M_2})] - 1}{\sqrt{\frac{R_1^2}{M_1 l^2} - 1}\sqrt{\frac{R_2^2}{M_2 l^2} - 1}}, \tag{B.75}$$

$$\beta_{mn}'^{(12)} = \frac{\frac{R_1 R_2}{\sqrt{M_1 M_2} l^2} \cosh[2\pi(m\sqrt{M_1} - n\sqrt{M_2})] + 1}{\sqrt{\frac{R_1^2}{M_1 l^2} - 1}\sqrt{\frac{R_2^2}{M_2 l^2} - 1}}. \tag{B.76}$$

Using these coordinates $\bar{t}$, we can obtain an analytical expression of $\widehat{W}_{BTZ}^{(12)}$, with the additional subtlety that for $\sqrt{M_1/M_2} \in \mathbb{Q}\backslash\{1\}$ and $R_1/\sqrt{M_1} = R_2/\sqrt{M_2}$ an additional pole arises, giving a singular term sing.:

$$\begin{aligned}\widehat{W}_{BTZ}^{(12)}(K) = \text{sing.} + \frac{1}{2\sqrt{\widetilde{\gamma_1}\widetilde{\gamma_2}}} \frac{1}{\sum_k \Upsilon^{2k}} \frac{1}{e^{K/T_H^{AdS}} + 1} \\ \sum_{m,n}\left[P_{\frac{iK}{2\pi T_H^{AdS}} - \frac{1}{2}}\left(\beta_{mn}^{(12)}\right) - \zeta P_{\frac{iK}{2\pi T_H^{AdS}} - \frac{1}{2}}\left(\beta_{mn}'^{(12)}\right)\right],\end{aligned} \tag{B.77}$$

with T_H^{AdS} the Hawking temperature of the AdS–Rindler horizon with respect to AdS–Rindler time, $T_H^{AdS} = 1/(2\pi l)$. If $\sqrt{M_1/M_2} \notin \mathbb{Q}\backslash\{1\}$, then we have sing. $= 0$,else it gives the following non-zero contribution if also $R_1/\sqrt{M_1} = R_2/\sqrt{M_2}$:

$$\text{sing.} = \frac{1}{4\sqrt{\widetilde{\gamma_1}\widetilde{\gamma_2}}pq} \quad \text{if } \sqrt{M_1/M_2} = p/q \in \mathbb{Q}\backslash\{1\} \text{ and } R_1/\sqrt{M_1} = R_2/\sqrt{M_2} \tag{B.78}$$

with $\gcd(p, q) = 1$.

B.6.2 Analytic expressions for single term

Consider the detector response function

$$\begin{aligned}\mathcal{F}(E) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{d}\tau \mathrm{d}\tau' \eta(\tau)\eta(\tau') e^{-iE(\tau-\tau')} W(\tau, \tau') \\ &= \gamma^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{d}t \mathrm{d}t' \eta(\gamma t)\eta(\gamma t') e^{-iE\gamma(t-t')} W(t - t')\end{aligned} \tag{B.79}$$

assuming $W(t, t') = W(t - t')$, with $\gamma = \gamma(R)$ with R the fixed radial coordinate of the static detector. Here and everywhere below W should be understood as W_{BTZ}. By applying the inverse Fourier transform to W, and Fourier transforming the switching functions η, η, we obtain

$$\mathcal{F}(E) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega |\widehat{\eta}(\tfrac{\omega}{\gamma})|^2 \widehat{W}(E\gamma + \omega), \tag{B.80}$$

with the Fourier transform defined as

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} ds f(s) \mathrm{e}^{-\mathrm{i}\omega s} \tag{B.81}$$

and inverse by

$$f(s) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{i\omega s} \hat{f}(\omega). \tag{B.82}$$

See Ref. [808] for a more precise treatment. For our Gaussian switching function, we have, for our detector at fixed location $r = R$,

$$|\widehat{\eta}(\tfrac{\omega}{\widetilde{\gamma}})|^2 = \frac{1}{N} 2\sigma^2 e^{-\frac{\omega^2\sigma^2}{\widetilde{\gamma}^2}} = \frac{1}{N} 2\sigma^2 e^{-\frac{\omega^2\sigma^2}{\widetilde{\gamma}^2 M}} \tag{B.83}$$

with N an appropriate normalization factor. If we choose σ to be very large, i.e. adiabatic switching, $\widehat{\eta}(\omega)$ becomes approximately a delta function, and we are basically probing $\widehat{\mathcal{W}}(E\gamma)$ for frequencies ω/γ. This $1/\gamma$ factor arising from the redshift is what leads to the local Tolman temperature $T(R)$ for a static detector at R being related to the Hawking temperature T_H by $T = T_H/\gamma(R)$ with $\gamma(R) = \sqrt{R^2/l^2 - M}$.

Now, to calculate $\widehat{\mathcal{W}}(K)$, we have,

$$\begin{aligned}
\widehat{W}(K) &= \int_{-\infty}^{+\infty} \mathrm{d}s e^{-iKs} W(s) \\
&= \frac{1}{4\pi l\sqrt{2}\widetilde{\gamma}} \sum_m \int_{-\infty}^{+\infty} \mathrm{d}s \Bigg[\frac{e^{-iKs}}{\sqrt{\alpha_m - \cosh\left(s\sqrt{M}/l - i\epsilon\right)}} \\
&\qquad - \frac{\zeta e^{-iKs}}{\sqrt{\alpha'_m - \cosh\left(s\sqrt{M}/l - i\epsilon\right)}} \Bigg] \\
&= \frac{1}{4\pi\sqrt{2}\gamma} \sum_m \int_{-\infty}^{+\infty} \mathrm{d}z \Bigg[\frac{e^{-iKlz/\sqrt{M}}}{\sqrt{\alpha_m - \cosh(z - i\epsilon)}} \\
&\qquad - \frac{\zeta e^{-iKlz/\sqrt{M}}}{\sqrt{\alpha'_m - \cosh(z - i\epsilon)}} \Bigg].
\end{aligned} \tag{B.84}$$

with $\widetilde{\gamma} = \gamma/\sqrt{M}$,

$$\alpha_m = \frac{\frac{R^2}{Ml^2}\cosh[2\pi m\sqrt{M}] - 1}{\frac{R^2}{Ml^2} - 1} = \frac{\frac{R^2}{Ml^2}\cosh[2\pi m\sqrt{M}] - 1}{\widetilde{\gamma}^2}. \tag{B.85}$$

and

$$\alpha'_m = \frac{\frac{R^2}{Ml^2}\cosh[2\pi m\sqrt{M}] + 1}{\frac{R^2}{Ml^2} - 1} = \frac{\frac{R^2}{Ml^2}\cosh[2\pi m\sqrt{M}] + 1}{\widetilde{\gamma}^2}, \tag{B.86}$$

the $z - i\epsilon$ prescription, and $\zeta = -1, 0, 1$ for Neumann, transparent or Dirichlet boundary conditions.

Note that we have $\alpha_0 = 1$ so that the term with α_0 has a pole at $z = 0$. In this case, we first rewrite using $\cosh(z) - 1 = 2\sinh^2(z/2)$, and using the Sokhotski formula

$$\frac{1}{\sinh(z - i\epsilon)} = i\pi\delta(z) + \mathrm{PV}\frac{1}{\sinh(z)}, \tag{B.87}$$

We find that the singular part gives rise to the term $1/(4l\widetilde{\gamma})$.

Following Ref. [746], the nonsingular parts (including the principal value part of the singular term with α_0 above) in the integrals in Eq. (B.84) can be performed, and give

$$\widehat{W}(K) = \frac{1}{4\gamma} + \frac{1}{2\gamma}\frac{1}{e^{K/T_H} + 1}\sum_n \left[P_{\frac{iK}{2\pi T_H} - \frac{1}{2}}(\alpha_n) - \zeta P_{\frac{iK}{2\pi T_H} - \frac{1}{2}}(\alpha_n')\right] \tag{B.88}$$

with $T_H = \sqrt{M}/(2\pi l)$ the Hawking temperature of the AdS–Rindler horizon with respect to AdS–Rindler time. Note that Ref. [746] left out the first term arising from the pole singularity.

B.6.3 Analytic expressions for interference term

For the interference term, we can obtain a similar expression. We now consider the general situation that comprises both the position- and mass-superposed black holes, with position and mass R_1, M_1 and R_2, M_2 in the two branches of the superposition, respectively. Consider the interference term in the detector transition probability

$$\mathcal{F}_{12}(E) = \int_{\tau_i}^{\tau_f} d\tau d\tau' \eta(\tau)\eta(\tau') e^{-iE(\tau - \tau')} W^{(12)}(x(\tau), x(\tau')), \tag{B.89}$$

with $W^{(12)}$ shorthand for $W^{(12)}_{BTZ}$. Using the AdS time coordinates $\bar{t} = t\sqrt{M_i} = \tau/\widetilde{\gamma}_i$ with $\widetilde{\gamma}_i = \sqrt{R_i^2/(M_i l^2) - 1}$ in the branch with BTZ black hole mass M_i, we find

$$\mathcal{F}_{12}(E) = \widetilde{\gamma_1}\widetilde{\gamma_2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{d}\bar{t}\mathrm{d}\bar{t}' \eta(\widetilde{\gamma_1}\bar{t})\eta(\widetilde{\gamma_2}\bar{t}') e^{-i(E\widetilde{\gamma_1}\bar{t} - E\widetilde{\gamma_2}\bar{t}')} W^{(12)}_{BTZ}(\bar{t} - \bar{t}'). \tag{B.90}$$

We will again try to write $\mathcal{F}_{12}$ in the form of Eq. (B.80). We start off by changing integration variables with $s = \bar{t} - \bar{t}'$ (we omit the integration limits here, all integrals,

also under the changes of variables that occur, run from $-\infty$ to $+\infty$):

$$\begin{aligned}
\mathcal{F}_{12}(E) &= \widetilde{\gamma_1}\widetilde{\gamma_2}\int\int \mathrm{d}s\mathrm{d}\bar{t}'\,\eta(\widetilde{\gamma_1}[s+\bar{t}'])\,\eta(\widetilde{\gamma_2}\bar{t}')e^{-iE\widetilde{\gamma_1}s}e^{-iE(\widetilde{\gamma_1}-\widetilde{\gamma_2})\bar{t}'}W^{(12)}_{BTZ}(s)\\
&= \widetilde{\gamma_1}\widetilde{\gamma_2}\int\int \mathrm{d}s\mathrm{d}\bar{t}'\,\eta(\widetilde{\gamma_1}[s+\bar{t}'])\,\eta(\widetilde{\gamma_2}\bar{t}')e^{-iE\widetilde{\gamma_1}s}e^{-iE(\widetilde{\gamma_1}-\widetilde{\gamma_2})\bar{t}'}\int\frac{\mathrm{d}\omega}{2\pi}\widehat{W}^{(12)}_{BTZ}(\omega)e^{i\omega s}\\
&= \frac{\widetilde{\gamma_1}\widetilde{\gamma_2}}{2\pi}\int\mathrm{d}\omega\int\int \mathrm{d}s\mathrm{d}\bar{t}'\,\eta(\widetilde{\gamma_1}[s+\bar{t}'])\,\eta(\widetilde{\gamma_2}\bar{t}')e^{i\omega s}e^{-iE(\widetilde{\gamma_1}-\widetilde{\gamma_2})\bar{t}'}\widehat{W}^{(12)}_{BTZ}(\omega+\widetilde{\gamma_1}E)\\
&= \frac{\widetilde{\gamma_2}}{2\pi}\int\mathrm{d}\omega\int\int \mathrm{d}\bar{t}'\mathrm{d}u\,\eta(u)e^{i\frac{\omega}{\widetilde{\gamma_1}}u}\eta(\widetilde{\gamma_2}\bar{t}')e^{-i[\omega+E(\widetilde{\gamma_1}-\widetilde{\gamma_2})]\bar{t}'}\widehat{W}^{(12)}_{BTZ}(\omega+\widetilde{\gamma_1}E)\\
&= \frac{1}{2\pi}\int\mathrm{d}\omega\int\mathrm{d}u\,\eta(u)e^{i\frac{\omega}{\widetilde{\gamma_1}}u}\int\mathrm{d}v\eta(v)e^{-i\left[\frac{\omega}{\widetilde{\gamma_2}}+E\left(\frac{\widetilde{\gamma_1}}{\widetilde{\gamma_2}}-1\right)\right]v}\widehat{W}^{(12)}_{BTZ}(\omega+\widetilde{\gamma_1}E)\\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{d}\omega\,\widehat{\eta}(\tfrac{\omega}{\widetilde{\gamma_1}})^*\,\widehat{\eta}(\tfrac{\omega}{\widetilde{\gamma_2}}+E[\tfrac{\widetilde{\gamma_1}}{\widetilde{\gamma_2}}-1])\,\widehat{W}^{(12)}_{BTZ}(\omega+\widetilde{\gamma_1}E)
\end{aligned} \tag{B.91}$$

where we changed variables to $u=\widetilde{\gamma_1}(s+\bar{t}'), v=\widetilde{\gamma_2}\bar{t}'$.

Next, the main calculation of this section, we calculate $\widehat{W}^{(12)}(K)$:

$$\begin{aligned}
\widehat{W}^{(12)}(K) &= \int_{-\infty}^{+\infty}\mathrm{d}s e^{-iKs}W^{(12)}(s)\\
&= \frac{1}{4\pi l\sqrt{2\widetilde{\gamma_1}\widetilde{\gamma_2}}}\frac{1}{\sum_k\Upsilon^{2k}}\sum_{m,n}\int_{-\infty}^{+\infty}\mathrm{d}s\left[\frac{e^{-iKs}}{\sqrt{\beta^{(12)}_{mn}-\cosh(s/l-i\epsilon)}}\right.\\
&\qquad\left.-\frac{\zeta e^{-iKs}}{\sqrt{\beta'^{(12)}_{mn}-\cosh(s/l-i\epsilon)}}\right]\\
&= \frac{1}{4\pi l\sqrt{2\widetilde{\gamma_1}\widetilde{\gamma_2}}}\frac{1}{\sum_k\Upsilon^{2k}}\sum_{m,n}\int_{-\infty}^{+\infty}\mathrm{d}z\left[\frac{e^{-iKlz}}{\sqrt{\beta^{(12)}_{mn}-\cosh(z-i\epsilon)}}\right.\\
&\qquad\left.-\frac{\zeta e^{-iKlz}}{\sqrt{\beta'^{(12)}_{mn}-\cosh(z-i\epsilon)}}\right].
\end{aligned} \tag{B.92}$$

with

$$\beta^{(12)}_{mn}=\frac{\frac{R_1R_2}{\sqrt{M_1M_2}l^2}\cosh[2\pi(m\sqrt{M_1}-n\sqrt{M_2})]-1}{\sqrt{\frac{R_1^2}{M_1l^2}-1}\sqrt{\frac{R_2^2}{M_2l^2}-1}} \tag{B.93}$$

and

$$\beta'^{(12)}_{mn}=\frac{\frac{R_1R_2}{\sqrt{M_1M_2}l^2}\cosh[2\pi(m\sqrt{M_1}-n\sqrt{M_2})]+1}{\sqrt{\frac{R_1^2}{M_1l^2}-1}\sqrt{\frac{R_2^2}{M_2l^2}-1}}, \tag{B.94}$$

and $\zeta=-1,0,1$ for Neumann, transparent or Dirichlet boundary conditions. Note that if, for some m,n we have that $\beta^{(12)}_{mn}=1$, which occurs at rational values of $\sqrt{M_1/M_2}$ in the

mass-superposed case (so $M_1 \neq M_2$) while $R_1/\sqrt{M_1} = R_2/\sqrt{M_2}$, then the term with $\beta^{(12)}_{mn}$ has a pole at $z = 0$. In this case, we first rewrite using $\cosh(z) - 1 = 2\sinh^2(z/2)$, and using the Sokhotski formula

$$\frac{1}{\sinh(z - i\epsilon)} = i\pi\delta(z) + \mathrm{PV}\frac{1}{\sinh(z)}, \tag{B.95}$$

we find the singular part giving rise to the term $1/(4l\sqrt{\widetilde{\gamma_1}\widetilde{\gamma_2}})$. We have to divide this term by $1/\sum_k \Upsilon^{2k}$, and it occurs for all $m = \sqrt{M_1/M_2}n$ (and $R_1/\sqrt{M_1} = R_2/\sqrt{M_2}$), This yields a finite contribution which in general might be cut-off dependent, i.e. depending on how the infinite sums are dealt with. However, there is a natural way to deal with the sum, arising from the group structure of the groups generated by the identification Γ_M that defines the BTZ spacetime as a quotient of AdS–Rindler (cf. Appendix B.2 and Appendix of Ref. [119]). Namely, let Γ_{M_i} denote the identification that defined the BTZ spacetime with mass M_i, $\Gamma^n_{M_i} : \varphi \mapsto \varphi + 2\pi n\sqrt{M_i}$. If $\sqrt{M_1/M_2} = p/q \in \mathbb{Q}$ with $\gcd(p, q) = 1$, then the groups G_1, G_2 generated by $\Gamma_{M_1}, \Gamma_{M_2}$ share a subgroup H, with indices $[G_1 : H] = q, [G_2 : H] = p$, suggesting their volumes relate as $\mathrm{Vol}(H)/\sqrt{\mathrm{Vol}(G_1)\mathrm{Vol}(G_2)} = 1/\sqrt{pq}$. The singularity is hit at every point of H (assuming also $R_1/\sqrt{M_1} = R_2/\sqrt{M_2}$), and a suggestive normalization of the BTZ two-point function is then

$$W^{(12)}_{\mathrm{BTZ}}(x, x') = \frac{1}{\sqrt{\mathrm{Vol}(G_1)\mathrm{Vol}(G_2)}} \sum_{\Gamma^n_{M_1} \in G_1} \sum_{\Gamma^m_{M_2} \in G_2} W_{\mathrm{AdS}}\left(\Gamma^n_{M_1} x, \Gamma^m_{M_2} x'\right) \tag{B.96}$$

for untwisted fields ($\Upsilon = 1$), arising from the identity for the automorphic field [119, 858]:

$$\phi_{\mathrm{BTZ},M_i}(x) := \frac{1}{\sqrt{\mathrm{Vol}(G_i)}} \sum_n \psi_{\text{AdS–Rindler}}\left(\Gamma^n_{M_i} x\right) \tag{B.97}$$

for untwisted fields. As the singular term gives a contribution $1/(4\sqrt{\widetilde{\gamma_1}\widetilde{\gamma_2}})$ at each element of H, we get a contribution

$$\frac{1}{4\sqrt{\widetilde{\gamma_1}\widetilde{\gamma_2}}} \frac{\mathrm{Vol}(H)}{\sqrt{\mathrm{Vol}(G_1)\mathrm{Vol}(G_2)}} = \frac{1}{4\sqrt{\widetilde{\gamma_1}\widetilde{\gamma_2}}} \frac{1}{\sqrt{pq}} \tag{B.98}$$

from the singularities. This thus gives the total contribution

$$\text{sing.} = \frac{1}{4\sqrt{\widetilde{\gamma_1}\widetilde{\gamma_2}pq}} \quad \text{for } \sqrt{M_1/M_2} = p/q \in \mathbb{Q}^+\backslash\{1\} \text{ and } R_1/\sqrt{M_1} = R_2/\sqrt{M_2} \tag{B.99}$$

with $\gcd(p, q) = 1$, to be inserted as the contribution of the (potential) singular part to $\widehat{W}^{(12)}(K)$. Had one taken a square-box regularisation of the sums instead (which is

perhaps less well motivated than the above volume argument), i.e., taking $|m|, |n|, |k| \leq N$ such that the double sum becomes $1/(2N+1)\sum_{|m|,|n|\leq N}$ one would get a contribution $1/(4\sqrt{\widetilde{\gamma_1}\widetilde{\gamma_2}}\max\{q,p\})$ for sing. above instead.

Regarding all other terms without poles, including the principal value part of the term that was singular before, following Ref. [746], the integrals in Eq. (B.92) can be performed, and give

$$\widehat{W}^{(12)}(K) = \text{sing.} + \frac{1}{2\sqrt{\widetilde{\gamma_1}\widetilde{\gamma_2}}}\frac{1}{\sum_k \Upsilon^{2k}}\frac{1}{e^{K/T_H^{AdS}}+1}\sum_{m,n}\Bigg[P_{\frac{iK}{2\pi T_H^{AdS}}-\frac{1}{2}}\left(\beta_{mn}^{(12)}\right) - \zeta P_{\frac{iK}{2\pi T_H^{AdS}}-\frac{1}{2}}\left(\beta_{mn}^{\prime(12)}\right)\Bigg], \tag{B.100}$$

with T_H^{AdS} the Hawking temperature of the AdS–Rindler horizon with respect to AdS–Rindler time, $T_H^{AdS} = 1/(2\pi l)$. If $\sqrt{M_1/M_2} \notin \mathbb{Q}\backslash\{1\}$, we have sing. $= 0$, else it gives the contribution as in Eq. (B.99).

In fact, for the position-superposition, we can remove the double sum in Eq. (B.100). Namely, for the position-superposed case we have $M_1 = M_2$ and we can write

$$\widehat{W}^{(12)}(K) = \frac{1}{2\sqrt{\widetilde{\gamma_1}\widetilde{\gamma_2}}}\frac{1}{e^{K/T_H^{AdS}}+1}\sum_{m}\left[P_{\frac{iK}{2\pi T_H^{AdS}}-\frac{1}{2}}\left(\alpha_m^{(12)}\right) - \zeta P_{\frac{iK}{2\pi T_H^{AdS}}-\frac{1}{2}}\left(\alpha_m^{\prime(12)}\right)\right], \tag{B.101}$$

with

$$\alpha_m^{(12)} = \frac{\frac{R_1R_2}{Ml^2}\cosh[2\pi m\sqrt{M}]-1}{\sqrt{\frac{R_1^2}{Ml^2}-1}\sqrt{\frac{R_2^2}{Ml^2}-1}}. \tag{B.102}$$

To be precise, in case $R_1 = R_2$, an additional pole contribution $1/(4\tilde{\gamma})$ to $\widehat{W}^{(12)}$ arises, but at $R_1 = R_2$ our initial 'superposed' state of Eq. (9.21) becomes an unnormalized non-superposed state, and similarly for the interferometric measurement effect, $|+\rangle_B = (|\mathbf{x}_1\rangle_B + |\mathbf{x}_2\rangle_B)/\sqrt{2}$.

A note about energy and time coordinates. We shortly comment upon how to compare energy in different frames, choosing different time coordinates, important for example for establishing when the detector's energy gap is small compared to the black hole mass concerning the approximation of neglecting the detector's gravitational backreaction, or for comparing energies across different spacetimes. Let $H_i^{(D)}$ be the detector's free Hamiltonian in its rest frame, in branch i of the superposition. We take the detector in each branch of the superposition to be identical, i.e. $H_i^{(D)} = H^{(D)}$. Then,

acting on the detector energy eigenstate $|E\rangle_D$ with energy eigenvalue $E^{(D)}$ we have $H|E\rangle_D = E^{(D)}|E\rangle_D$ in the detector frame. If we consider the detector in a superposition of locations, or in a superposed spacetime, with branches labelled by i, we need to be able to relate this detector frame energy to energy and Hamiltonians with respect to other time parametrizations. The detector proper time and coordinate time are related by $\tau_i = \gamma_i t_i$ in the branch labelled by i. In such branch, the Hamiltonians are related by $H^{(D)} = H^{\tau_i} = dt/d\tau H^{t_i} = H^{t_i}/\gamma_i$. Therefore, the energies are related by

$$E^{(D)} = E^{\tau,i} = \frac{1}{\gamma_i(R_i)} E^{t_i}. \tag{B.103}$$

Moving to AdS–Rindler coordinates, we find

$$E^{(D)} = E^{\tau,i} = \frac{1}{\widetilde{\gamma}_i(R_i)} E^{\bar{t}_i}. \tag{B.104}$$

References

[1] J. S. Bell. "On the Impossible Pilot Wave". *Found. Phys.* 12 (1982), pp. 989–999. DOI: 10.1007/BF01889272.

[2] T. Maudlin. "Three Measurement Problems". *Topoi* 14.1 (1995), pp. 7–15. DOI: 10.1007/bf00763473.

[3] I. Pitowsky. "Quantum Mechanics as a Theory of Probability". *Festschrift in Honor of Jeffrey Bub*. Ed. by W. Demopoulos and I. Pitowsky. Western Ontario Series in Philosophy of Science. New York: Springer, 2007. DOI: 10.48550/arXiv.quant-ph/0510095.

[4] J. Bub and I. Pitowsky. "Two Dogmas About Quantum Mechanics". *Many Worlds? Everett, Quantum Theory, and Reality*. Ed. by S. Saunders, J. Barrett, A. Kent, and D. Wallace. Oxford University Press, 2010, pp. 431–456. DOI: 10.48550/arXiv.0712.4258.

[5] Č. Brukner. "On the quantum measurement problem". *Quantum [Un] Speakables II: Half a Century of Bell's Theorem* (2017), pp. 95–117. DOI: 10.48550/arXiv.1507.05255.

[6] M. Schlosshauer. "Decoherence, the measurement problem, and interpretations of quantum mechanics". *Rev. Mod. Phys.* 76 (4 2005), pp. 1267–1305. DOI: 10.1103/RevModPhys.76.1267.

[7] A. J. Leggett. "The Quantum Measurement Problem". *Science* 307.5711 (2005), pp. 871–872. DOI: 10.1126/science.1109541. eprint: https://www.science.org/doi/pdf/10.1126/science.1109541.

[8] J. R. Hance and S. Hossenfelder. "What does it take to solve the measurement problem?" *J. Phys. Commun.* 6.10 (2022), p. 102001. ISSN: 2399-6528. DOI: 10.1088/2399-6528/ac96cf.

[9] G. A. D. Briggs, J. N. Butterfield, and A. Zeilinger. "The Oxford Questions on the foundations of quantum physics". *Proc. Roy. Soc. Lond. A* 469 (2013), p. 20130299. DOI: 10.1098/rspa.2013.0299. arXiv: 1307.1310 [quant-ph].

[10] E. P. Wigner. "Remarks on the mind-body question". *Philosophical reflections and syntheses* (1995), pp. 247–260. DOI: 10.1007/978-3-642-78374-6_20.

[11] D. Frauchiger and R. Renner. "Quantum theory cannot consistently describe the use of itself". *Nat. Commun.* 9.1 (2018), p. 3711. DOI: 10.1038/s41467-018-05739-8.

[12] K.-W. Bong, A. Utreras-Alarcón, F. Ghafari, Y.-C. Liang, N. Tischler, E. G. Cavalcanti, G. J. Pryde, and H. M. Wiseman. "A strong no-go theorem on the Wigner's friend paradox". *Nat. Phys.* 16.12 (2020), pp. 1199–1205. DOI: 10.1038/s41567-020-0990-x.

[13] E. G. Cavalcanti and H. M. Wiseman. "Implications of Local Friendliness violation for quantum causality". *Entropy* 23.8 (2021), p. 925. DOI: 10.3390/e23080925.

[14] M. Haddara and E. G. Cavalcanti. "A possibilistic no-go theorem on the Wigner's friend paradox". *New. J. Phys.* 25.9 (2023), p. 093028. DOI: 10.1088/1367-2630/aceea3.

[15] P. A. Guérin, V. Baumann, F. Del Santo, and Č. Brukner. "A no-go theorem for the persistent reality of Wigner's friend's perception". *Commun. Phys.* 4.1 (2021), p. 93. DOI: 10.1038/s42005-021-00589-1.

[16] A. Utreras-Alarcón, E. G. Cavalcanti, and H. M. Wiseman. *Allowing Wigner's friend to sequentially measure incompatible observables*. 2023. arXiv: 2305.09102 [quant-ph].

[17] M. Haddara and E. G. Cavalcanti. *Local Friendliness Polytopes In Multipartite Scenarios*. 2024. arXiv: 2407.20346 [quant-ph].

[18] J. Szangolies. *The Quantum Rashomon Effect: A Strengthened Frauchiger-Renner Argument*. 2023. arXiv: 2011.12716 [quant-ph].

[19] L. Walleghem, R. S. Barbosa, M. Pusey, and S. Weigert. *A refined Frauchiger–Renner paradox based on strong contextuality*. 2026. DOI: 10.22331/q-2026-05-26-2116.

[20] L. Walleghem, R. Wagner, Y. Yīng, and D. Schmid. "Extended Wigner's friend paradoxes do not require nonlocal correlations". *Phys. Rev. A* 112.2 (2025). ISSN: 2469-9934. DOI: 10.1103/n4hv-rlgj.

[21] L. Walleghem, Y. Yīng, R. Wagner, and D. Schmid. *Connecting extended Wigner's friend arguments and noncontextuality*. 2025. DOI: 10.22331/q-2025-07-31-1819.

[22] L. Walleghem and L. Catani. *An extended Wigner's friend no-go theorem inspired by generalized contextuality*. 2025. arXiv: 2502.02461 [quant-ph].

[23] V. Vilasini, N. Nurgalieva, and L. del Rio. "Multi-agent paradoxes beyond quantum theory". *New. J. Phys.* 21.11 (2019), p. 113028. DOI: 10.1088/1367-2630/ab4fc4.

[24] V. Vilasini and M. P. Woods. *A general framework for consistent logical reasoning in Wigner's friend scenarios: subjective perspectives of agents within a single quantum circuit*. 2022. arXiv: 2209.09281 [quant-ph].

[25] N. Ormrod, V. Vilasini, and J. Barrett. *Which theories have a measurement problem?* 2023. arXiv: 2303.03353 [quant-ph].

[26] G. Leegwater. "When Greenberger, Horne and Zeilinger meet Wigner's Friend". *Found. Phys.* 52.4 (2022), p. 68. DOI: 10.1007/s10701-022-00586-6.

[27] D. Schmid, Y. Yīng, and M. Leifer. *A review and analysis of six extended Wigner's friend arguments*. 2023. DOI: 10.48550/arXiv.2308.16220. arXiv: 2308.16220 [quant-ph].

[28] Č. Brukner. "A no-go theorem for observer-independent facts". *Entropy* 20.5 (2018), p. 350. DOI: 10.3390/e20050350.

[29] N. Ormrod and J. Barrett. *A no-go theorem for absolute observed events without inequalities or modal logic*. 2022. arXiv: 2209.03940 [quant-ph].

[30] N. Nurgalieva and L. del Rio. *Inadequacy of modal logic in quantum settings*. 2018. arXiv: 1804.01106 [quant-ph].

[31] S. B. Montanhano. *Contextuality in multi-agent paradoxes*. 2023. arXiv: 2305.07792 [quant-ph].

[32] M. Żukowski and M. Markiewicz. "Physics and metaphysics of Wigner's friends: Even performed premeasurements have no results". *Phys. Rev. Lett.* 126.13 (2021), p. 130402. DOI: 10.1103/PhysRevLett.126.130402.

[33] Y. Yīng, M. M. Ansanelli, A. D. Biagio, E. Wolfe, D. Schmid, and E. G. Cavalcanti. "Relating Wigner's Friend Scenarios to Nonclassical Causal Compatibility, Monogamy Relations, and Fine Tuning". *Quantum* 8 (2024), p. 1485. ISSN: 2521-327X. DOI: 10.22331/q-2024-09-26-1485.

[34] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. "Bell nonlocality". *Rev. Mod. Phys.* 86 (2 2014), pp. 419–478. DOI: 10.1103/RevModPhys.86.419.

[35] J. S. Bell. "On the Einstein Podolsky Rosen paradox". *Physics Physique Fizika* 1.3 (1964), p. 195. DOI: 10.1142/9789812386540_0002.

[36] J. S. Bell. "The Theory of Local Beables". *Epistemological Letters* 9 (1976), pp. 11–24.

[37] A. Einstein, M. Born, and H. Born. *The Born-Einstein Letters: The Correspondence between Max & Hedwig Born and Albert Einstein 1916–1955*. 1st ed. London and Basingstoke: The MacMillan Press, 1971.

[38] M. F. Pusey, J. Barrett, and T. Rudolph. "On the reality of the quantum state". *Nat. Phys.* 8.6 (2012), pp. 475–478. DOI: 10.1038/nphys2309.

[39] A. Peres. "Unperformed experiments have no results". *Am. J. Phys.* 46.7 (1978), pp. 745–747. DOI: 10.1119/1.11393.

[40] S. Kochen and E. P. Specker. "The Problem of Hidden Variables in Quantum Mechanics". *Ernst Specker Selecta*. Birkhäuser Basel, 1990, pp. 235–263. DOI: 10.1007/978-3-0348-9259-9_21.

[41] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J.-Å. Larsson. "Kochen-Specker contextuality". *Rev. Mod. Phys.* 94 (4 2022), p. 045007. DOI: 10.1103/RevModPhys.94.045007.

[42] S. Abramsky and A. Brandenburger. "The sheaf-theoretic structure of non-locality and contextuality". *New. J. Phys.* 13.11 (2011), p. 113036. DOI: 10.1088/1367-2630/13/11/113036.

[43] R. W. Spekkens. "Contextuality for preparations, transformations, and unsharp measurements". *Phys. Rev. A* 71.5 (2005), p. 052108. DOI: 10.1103/PhysRevA.71.052108.

[44] D. Schmid, J. H. Selby, M. F. Pusey, and R. W. Spekkens. "A structure theorem for generalized-noncontextual ontological models". *Quantum* 8 (2024), p. 1283. DOI: 10.22331/q-2024-03-14-1283.

[45] S. W. Hawking. "Particle creation by black holes". *Commun. Math. Phys.* 43.3 (1975), pp. 199–220. DOI: 10.1007/BF02345020.

[46] S. W. Hawking. "Breakdown of predictability in gravitational collapse". *Phys. Rev. D* 14.10 (1976), p. 2460. DOI: 10.1103/PhysRevD.14.2460.

[47] L. Susskind and L. Thorlacius. "Gedanken experiments involving black holes". *Phys. Rev. D* 49 (2 1994), pp. 966–974. DOI: 10.1103/PhysRevD.49.966.

[48] P. Hayden and J. Preskill. "Black holes as mirrors: quantum information in random subsystems". *J. High Energy Phys.* 2007.09 (2007), pp. 120–120. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2007/09/120.

[49] D. Harlow. "Jerusalem lectures on black holes and quantum information". *Rev. Modern Phys.* 88.1 (2016), p. 015002. DOI: 10.1103/RevModPhys.88.015002.

[50] S. L. Braunstein and A. K. Pati. "Quantum information cannot be completely hidden in correlations: implications for the black-hole information paradox". *Phys. Rev. Lett.* 98.8 (2007), p. 080502. DOI: 10.1103/PhysRevLett.98.080502.

[51] S. L. Braunstein, S. Pirandola, and K. Życzkowski. "Better late than never: information retrieval from black holes". *Phys. Rev. Lett.* 110.10 (2013), p. 101301. DOI: 10.1103/PhysRevLett.110.101301.

[52] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully. "Black holes: complementarity or firewalls?" *J. High Energy Phys.* 2013.2 (2013), pp. 1–20. DOI: 10.1007/JHEP02%282013%29062.

[53] S. Aaronson. *The Complexity of Quantum States and Transformations: From Quantum Money to Black Holes*. 2016. arXiv: 1607.05256 [quant-ph].

[54] W. K. Wootters and W. H. Zurek. "A single quantum cannot be cloned". *Nature* 299.5886 (1982), pp. 802–803. DOI: 10.1038/299802a0.

[55] D. Dieks. "Communication by EPR devices". *Phys. Lett.A* 92.6 (1982), pp. 271–272. ISSN: 0375-9601. DOI: 10.1016/0375-9601(82)90084-6.

[56] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher. "Noncommuting Mixed States Cannot Be Broadcast". *Phys. Rev. Lett.* 76 (15 1996), pp. 2818–2821. DOI: 10.1103/PhysRevLett.76.2818.

[57] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[58] V. Coffman, J. Kundu, and W. K. Wootters. "Distributed entanglement". *Phys. Rev. A* 61 (5 2000), p. 052306. DOI: 10.1103/PhysRevA.61.052306.

[59] T. J. Osborne and F. Verstraete. "General Monogamy Inequality for Bipartite Qubit Entanglement". *Phys. Rev. Lett.* 96.22 (2006), p. 220503. DOI: 10.1103/PhysRevLett.96.220503. arXiv: quant-ph/0502176.

[60] M. Koashi and A. Winter. "Monogamy of quantum entanglement and other correlations". *Phys. Rev. A* 69.2 (2004), p. 022309. DOI: 10.1103/PhysRevA.69.022309. arXiv: quant-ph/0310037.

[61] N. Friis, G. Vitagliano, M. Malik, and M. Huber. "Entanglement certification from theory to experiment". *Nature Rev. Phys.* 1.1 (2018), pp. 72–87. DOI: 10.1038/s42254-018-0003-5. arXiv: 1906.10929 [quant-ph].

[62] D. N. Page. "Information in black hole radiation". *Phys. Rev. Lett.* 71.23 (1993), p. 3743. DOI: 10.1103/PhysRevLett.71.3743.

[63] D. N. Page. "Time dependence of Hawking radiation entropy". *J. Cosmol. Astropart. Phys.* 2013.09 (2013), p. 028. DOI: 10.1088/1475-7516/2013/09/028.

[64] A. Laddha, S. Prabhu, S. Raju, and P. Shrivastava. "The Holographic Nature of Null Infinity". *Scipost Phys.* 10 (2021), p. 041. ISSN: 2542-4653. DOI: 10.21468/scipostphys.10.2.041.

[65] H. Geng, A. Karch, C. Perez-Pardavila, S. Raju, L. Randall, M. Riojas, and S. Shashi. "Inconsistency of islands in theories with long-range gravity". *J. High Energy Phys.* 2022.1 (2022), pp. 1–46. DOI: 10.1007/jhep01(2022)182.

[66] C. F. Uhlemann. "Information transfer with a twist". *J. High Energy Phys.* 2022.1 (2022). ISSN: 1029-8479. DOI: 10.1007/jhep01(2022)126.

[67] C. F. Uhlemann. "Islands and Page curves in 4d from Type IIB". *J. High Energy Phys.* 2021.8 (2021). ISSN: 1029-8479. DOI: 10.1007/jhep08(2021)104.

[68] H. Geng, A. Karch, C. Perez-Pardavila, S. Raju, L. Randall, M. Riojas, and S. Shashi. "Information transfer with a gravitating bath". *SciPost Phys.* 10 (2021), p. 103. DOI: 10.21468/SciPostPhys.10.5.103.

[69] T. Banks, L. Susskind, and M. E. Peskin. "Difficulties for the Evolution of Pure States Into Mixed States". *Nucl. Phys. B* 244 (1984), pp. 125–134. DOI: 10.1016/0550-3213(84)90184-6.

[70] D. N. Page. "Is Black-Hole Evaporation Predictable?" *Phys. Rev. Lett.* 44 (5 1980), pp. 301–304. DOI: 10.1103/PhysRevLett.44.301.

[71] G. 't Hooft. "The black hole interpretation of string theory". *Nucl. Phys. B* 335 (1990), pp. 138–154. DOI: 10.1016/0550-3213(90)90174-C.

[72] L. Susskind, L. Thorlacius, and J. Uglum. "The stretched horizon and black hole complementarity". *Phys. Rev. D* 48 (8 1993), pp. 3743–3761. DOI: 10.1103/PhysRevD.48.3743.

[73] Y. Kiem, H. Verlinde, and E. Verlinde. "Black hole horizons and complementarity". *Phys. Rev. D* 52 (12 1995), pp. 7053–7065. DOI: 10.1103/PhysRevD.52.7053.

[74] J. Maldacena. "The large-N limit of superconformal field theories and supergravity". *Int. J. Theor. Phys.* 38.4 (1999), pp. 1113–1133. DOI: 10.1023/A%3A1026654312961.

[75] G. Penington. "Entanglement wedge reconstruction and the information paradox". *J. High Energy Phys.* 2020.9 (2020), pp. 1–84. DOI: 10.48550/arXiv.1905.08255.

[76] A. Almheiri, R. Mahajan, J. Maldacena, and Y. Zhao. "The Page curve of Hawking radiation from semiclassical geometry". *J. High Energy Phys.* 2020.3 (2020), pp. 1–24. DOI: 10.1007/jhep03(2020)149.

[77] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield. "The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole". *J. High Energy Phys.* 2019.12 (2019), pp. 1–47. DOI: 10.1007/jhep12(2019)063.

[78] L. Wang and R. Li. "Entanglement islands and the Page curve of Hawking radiation for rotating Kerr black holes". *Phys. Rev. D* 110 (6 2024), p. 066012. DOI: 10.1103/PhysRevD.110.066012.

[79] T. Hartman, E. Shaghoulian, and A. Strominger. "Islands in asymptotically flat 2D gravity". *J. High Energy Phys.* 2020.7 (2020). ISSN: 1029-8479. DOI: 10.1007/jhep07(2020)022.

[80] X. Wang, R. Li, and J. Wang. "Page curves for a family of exactly solvable evaporating black holes". *Phys. Rev. D* 103 (12 2021), p. 126026. DOI: 10.1103/PhysRevD.103.126026.

[81] V. Balasubramanian, A. Lawrence, J. M. Magán, and M. Sasieta. "Microscopic Origin of the Entropy of Black Holes in General Relativity". *Phys. Rev. X* 14 (1 2024), p. 011024. DOI: 10.1103/PhysRevX.14.011024.

[82] K. Hashimoto, N. Iizuka, and Y. Matsuo. "Islands in Schwarzschild black holes". *J. High Energy Phys.* 2020.6 (2020). ISSN: 1029-8479. DOI: 10.1007/jhep06(2020)085.

[83] V. Balasubramanian, A. Kar, and T. Ugajin. "Islands in de Sitter space". *J. High Energy Phys.* 2021.2 (2021). ISSN: 1029-8479. DOI: 10.1007/jhep02(2021)072.

[84] A. Almheiri, R. Mahajan, and J. E. Santos. "Entanglement islands in higher dimensions". *SciPost Phys.* 9 (2020), p. 001. DOI: 10.21468/SciPostPhys.9.1.001.

[85] T. Kibe, P. Mandayam, and A. Mukhopadhyay. "Holographic spacetime, black holes and quantum error correcting codes: a review". *The European Physical Journal C* 82.5 (2022), p. 463. DOI: 10.1140/epjc/s10052-022-10382-1.

[86] H. Z. Chen, R. C. Myers, D. Neuenfeld, I. A. Reyes, and J. Sandor. *Quantum Extremal Islands Made Easy, Part I: Entanglement on the Brane*. 2020. arXiv: 2006.04851 [hep-th].

[87] H. Z. Chen, R. C. Myers, D. Neuenfeld, I. A. Reyes, and J. Sandor. *Quantum Extremal Islands Made Easy, Part II: Black Holes on the Brane*. 2020. arXiv: 2010.00018 [hep-th].

[88] R. Bousso and G. Penington. *Islands Far Outside the Horizon*. arXiv:2312.03078 [hep-th]. 2023. DOI: 10.48550/arXiv.2312.03078.

[89] A. Almheiri, R. Mahajan, and J. Maldacena. *Islands outside the horizon*. 2023. arXiv: 1910.11077 [hep-th].

[90] T. Hartman, Y. Jiang, and E. Shaghoulian. "Islands in cosmology". *J. High Energy Phys.* 2020.11 (2020). ISSN: 1029-8479. DOI: 10.1007/jhep11(2020)111.

[91] C. Krishnan. "Critical islands". *J. High Energy Phys.* 2021.1 (2021). ISSN: 1029-8479. DOI: 10.1007/jhep01(2021)179.

[92] M. Van Raamsdonk. "Comments on wormholes, ensembles, and cosmology". *J. High Energy Phys.* 2021.12 (2021). ISSN: 1029-8479. DOI: 10.1007/jhep12(2021)156.

[93] H. Geng, Y. Nomura, and H.-Y. Sun. "Information paradox and its resolution in de Sitter holography". *Phys. Rev. D* 103 (12 2021), p. 126004. DOI: 10.1103/PhysRevD.103.126004.

[94] H. Geng, A. Karch, C. Perez-Pardavila, S. Raju, L. Randall, and M. Riojas. *Seeing Page Curves and Islands with Blinders On*. 2026. arXiv: 2602.06543 [hep-th].

[95] H. Geng and A. Karch. "Massive islands". *J. High Energy Phys.* 2020.9 (2020). ISSN: 1029-8479. DOI: 10.1007/jhep09(2020)121.

[96] H. Geng. *Revisiting Recent Progress in the Karch-Randall Braneworld*. 2025. arXiv: 2306.15671 [hep-th].

[97] S. Antonini, C.-H. Chen, H. Maxfield, and G. Penington. *An apologia for islands*. 2025. arXiv: 2506.04311 [hep-th].

[98] S. Raju. "Failure of the split property in gravity and the information paradox". *Class. Quantum Grav.* 39.6 (2022), p. 064002. DOI: 10.1088/1361-6382/ac482b.

[99] H. Geng. *Making the Case for Massive Islands*. 2025. arXiv: 2509.22775.

[100] H. Geng. *The Mechanism behind the Information Encoding for Islands*. 2026. arXiv: 2502.08703 [hep-th].

[101] S. W. Hawking, M. J. Perry, and A. Strominger. "Soft hair on black holes". *Phys. Rev. Lett.* 116.23 (2016), p. 231301. DOI: 10.1103/physrevlett.116.231301.

[102] A. Strominger. *Lectures on the Infrared Structure of Gravity and Gauge Theory*. 2018. arXiv: 1703.05448 [hep-th].

[103] D. Carney, L. Chaurette, D. Neuenfeld, and G. W. Semenoff. "Infrared quantum information". *Phys. Rev. Lett.* 119.18 (2017). arXiv:1706.03782 [hep-th], p. 180502. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.119.180502.

[104] D. Carney, L. Chaurette, D. Neuenfeld, and G. W. Semenoff. "Dressed infrared quantum information". *Phys. Rev.D* 97.2 (2018), p. 025007. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.97.025007.

[105] A. Strominger. *Black Hole Information Revisited*. 2017. arXiv: 1706.07143 [hep-th].

[106] E. E. Flanagan. "Order-Unity Correction to Hawking Radiation". *Phys. Rev. Lett.* 127.4 (2021), p. 041301. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.127.041301.

[107] L. Hausmann and R. Renner. *The firewall paradox is Wigner's friend paradox*. 2025. arXiv: 2504.03835 [quant-ph].

[108] M. Bañados, M. Henneaux, C. Teitelboim, and J. Zanelli. "Geometry of the 2+1 black hole". *Phys. Rev.D* 48.4 (1993), pp. 1506–1525. ISSN: 0556-2821. DOI: 10.1103/PhysRevD.48.1506.

[109] S. Carlip. "The (2 + 1)-dimensional black hole". *Class. Quantum Grav.* 12.12 (1995), pp. 2853–2879. ISSN: 1361-6382. DOI: 10.1088/0264-9381/12/12/005.

[110] G. Compère, R. Oliveri, and A. Seraj. "The Poincaré and BMS flux-balance laws with application to binary systems". *J. High Energy Phys.* 2020.10 (2020), p. 116. ISSN: 1029-8479. DOI: 10.1007/JHEP10(2020)116.

[111] E. E. Flanagan and D. A. Nichols. "Conserved charges of the extended Bondi-Metzner-Sachs algebra". *Phys. Rev. D* 95.4 (2017), p. 044002. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.95.044002.

[112] E. E. Flanagan. "Infrared Effects in the Late Stages of Black Hole Evaporation". *J. High Energy Phys.* 2021.7 (2021). arXiv:2102.13629 [hep-th], p. 137. ISSN: 1029-8479. DOI: 10.1007/JHEP07(2021)137.

[113] N. Bao, S. M. Carroll, A. Chatwin-Davies, J. Pollack, and G. N. Remmen. "Branches of the black hole wave function need not contain firewalls". *Phys. Rev. D* 97 (2018), p. 126014. ISSN: 2470-0029. DOI: 10.1103/physrevd.97.126014.

[114] X. Calmet and S. D. Hsu. "A brief history of Hawking's information paradox". *Europhys. Lett.* 139 (2022), p. 49001. DOI: 10.1209/0295-5075/ac81e8.

[115] A. Akil, L. Giannelli, L. Modesto, O. Dahlsten, G. Chiribella, and Č. Brukner. *A Quantum Superposition of Black Hole Evaporation Histories: Recovering Unitarity*. 2025. arXiv: 2507.17031 [gr-qc].

[116] S. Pasterski and H. Verlinde. "HPS meets AMPS: How Soft Hair Dissolves the Firewall". *J. High Energy Phys.* 2021.9 (2021), p. 99. ISSN: 1029-8479. DOI: 10.1007/JHEP09(2021)099.

[117] J. Hutchinson and D. Stojkovic. "Icezones instead of firewalls: extended entanglement beyond the event horizon and unitary evaporation of a black hole". *Class. Quantum Grav.* 33 (2016), p. 135006. ISSN: 1361-6382. DOI: 10.1088/0264-9381/33/13/135006.

[118] A. Strominger and A. Zhiboedov. "Gravitational memory, BMS supertranslations and soft theorems". *J. High Energy Phys.* 2016.1 (2016), p. 86. ISSN: 1029-8479. DOI: 10.1007/JHEP01(2016)086.

[119] J. Foo, C. S. Arabaci, M. Zych, and R. B. Mann. "Quantum signatures of black hole mass superpositions". *Phys. Rev. Lett.* 129.18 (2022), p. 181301. DOI: 10.1103/PhysRevLett.129.181301.

[120] L. Walleghem. *Stunned by Sleeping Beauty: How Prince Probability updates his forecast upon their fateful encounter*. 2024. arXiv: 2408.06797 [math.PR].

[121] L. Walleghem. *Are free choices absolute, when internalized in Wigner's friend?* 2026. arXiv: 2605.14538 [quant-ph].

[122] V. Moretti. *Spectral theory and quantum mechanics*. Springer, 2017. DOI: 10.1007/978-88-470-2835-7.

[123] B. C. Hall. *Quantum theory for mathematicians*. Springer, 2013. DOI: 10.1007/978-1-4614-7116-5.

[124] K. Landsman. *Foundations of quantum theory: From classical concepts to operator algebras*. Springer, 2017. DOI: 10.1007/978-3-319-51777-3.

[125] R. Haag. *Local quantum physics: Fields, particles, algebras*. Springer Berlin, Heidelberg, 1996. DOI: 10.1007/978-3-642-61458-3.

[126] R. M. Wald. *Quantum Field Theory in Curved Space-Time and Black Hole Thermodynamics*. Chicago Lectures in Physics. Chicago, IL: University of Chicago Press, 1995. ISBN: 978-0-226-87027-4.

[127] C. J. Fewster and R. Verch. *Measurement in Quantum Field Theory*. 2024. arXiv: 2304.13356 [math-ph].

[128] R. Renner. *Lecture Notes: Quantum Information Theory*. Lecture notes. 2020.

[129] M. M. Wilde. *Quantum Information Theory*. 2nd ed. Cambridge University Press, 2017. DOI: doi.org/10.1017/9781316809976.

[130] S. Weinberg. *The Quantum Theory of Fields*. Cambridge University Press, 1995. DOI: 10.1017/CBO9781139644167.

[131] J. A. Wheeler. "Superspace and the nature of quantum geometrodynamics". *Battelle Rencontres* (1968), pp. 242–307. DOI: 10.1007/978-3-642-88504-47.

[132] B. S. DeWitt. "Quantum Theory of Gravity. I. The Canonical Theory". *Phys. Rev.* 160 (5 1967), pp. 1113–1148. DOI: 10.1103/PhysRev.160.1113.

[133] D. N. Page and W. K. Wootters. "Evolution without evolution: Dynamics described by stationary observables". *Phys. Rev. D* 27 (12 1983), pp. 2885–2892. DOI: 10.1103/PhysRevD.27.2885.

[134] W. K. Wootters. ""Time" replaced by quantum correlations". *Int. J. Theor. Phys.* 23 (1984), pp. 701–711. DOI: https://doi.org/10.1007/BF02214098.

[135] V. Giovannetti, S. Lloyd, and L. Maccone. "Quantum time". *Phys. Rev. D* 92 (4 2015), p. 045033. DOI: 10.1103/PhysRevD.92.045033.

[136] A. R. H. Smith and M. Ahmadi. "Quantizing time: Interacting clocks and systems". *Quantum* 3 (2019), p. 160. ISSN: 2521-327X. DOI: 10.22331/q-2019-07-08-160.

[137] A. Boette and R. Rossignoli. "History states of systems and operators". *Phys. Rev. A* 98 (3 2018), p. 032108. DOI: 10.1103/PhysRevA.98.032108.

[138] R. Gambini, R. A. Porto, and J. Pullin. "Fundamental decoherence from quantum gravity: a pedagogical review". *Gen. Relativ. Gravit.* 39 (2007), pp. 1143–1156. DOI: https://doi.org/10.1007/s10714-007-0451-1.

[139] V. Baumann, F. D. Santo, A. R. H. Smith, F. Giacomini, E. Castro-Ruiz, and C. Brukner. "Generalized probability rules from a timeless formulation of Wigner's friend scenarios". *Quantum* 5 (2021), p. 524. ISSN: 2521-327X. DOI: 10.22331/q-2021-08-16-524.

[140] C. Marletto and V. Vedral. "Evolution without evolution and without ambiguities". *Phys. Rev. D* 95 (4 2017), p. 043510. DOI: 10.1103/PhysRevD.95.043510.

[141] A. Nikolova, G. K. Brennen, T. J. Osborne, G. J. Milburn, and T. M. Stace. "Relational time in anyonic systems". *Phys. Rev. A* 97 (3 2018), p. 030101. DOI: 10.1103/PhysRevA.97.030101.

[142] N. L. Diaz, J. M. Matera, and R. Rossignoli. "History state formalism for scalar particles". *Phys. Rev. D* 100 (12 2019), p. 125020. DOI: 10.1103/PhysRevD.100.125020.

[143] E. Castro-Ruiz, F. Giacomini, A. Belenchia, and Č. Brukner. "Quantum clocks and the temporal localisability of events in the presence of gravitating quantum systems". *Nat. Commun.* 11.1 (2020), p. 2672. DOI: https://doi.org/10.1038/s41467-020-16013-1.

[144] A. R. Smith and M. Ahmadi. "Quantum clocks observe classical and quantum time dilation". *Nat. Commun.* 11.1 (2020), p. 5360. DOI: https://doi.org/10.1038/s41467-020-18264-4.

[145] R. Gambini and J. Pullin. "Relational physics with real rods and clocks and the measurement problem of quantum mechanics". *Found. Phys.* (2007), pp. 1074–1092. DOI: https://doi.org/10.1007/s10701-007-9144-6.

[146] P. A. Höhn and A. Vanrietvelde. "How to switch between relational quantum clocks". *New J. Phys.* 22.12 (2020), p. 123048. DOI: 10.1088/1367-2630/abd1ac.

[147] E. Castro Ruiz, F. Giacomini, and Č. Brukner. "Entanglement of quantum clocks through gravity". *Proc. Natl. Acad. Sci.* 114.12 (2017). ISSN: 1091-6490. DOI: 10.1073/pnas.1616427114.

[148] S. Gogioso. "A diagrammatic approach to quantum dynamics". Vol. 139. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019, 19:1–19:23. DOI: 10.4230/LIPICS.CALCO.2019.19.

[149] A. Blommaert, T. G. Mertens, and H. Verschelde. "Clocks and rods in Jackiw-Teitelboim quantum gravity". *J. High Energy Phys.* 2019 (2019), p. 60. ISSN: 1029-8479. DOI: 10.1007/jhep09(2019)060.

[150] H. Geng. *Quantum Rods and Clock in a Gravitational Universe*. 2025. arXiv: 2412.03636 [hep-th].

[151] F. Giacomini. "Spacetime Quantum Reference Frames and superpositions of proper times". *Quantum* 5 (2021), p. 508. ISSN: 2521-327X. DOI: 10.22331/q-2021-07-22-508.

[152] H. Wang, F. Giacomini, F. Nori, and M. P. Blencowe. "Relational superposition measurements with a material quantum ruler". *Quantum* 8 (2024), p. 1335.

[153] L. Hardy. *Quantum Theory From Five Reasonable Axioms*. 2001. arXiv: quant-ph/0101012 [quant-ph].

[154] L. Hardy. *Reformulating and Reconstructing Quantum Theory*. 2011. arXiv: 1104.2066 [quant-ph].

[155] L. Hardy. *Reconstructing quantum theory*. 2013. arXiv: 1303.1538 [quant-ph].

[156] G. Chiribella, G. M. D'Ariano, and P. Perinotti. "Probabilistic theories with purification". *Phys. Rev. A* 81.6 (2010). ISSN: 1094-1622. DOI: 10.1103/physreva.81.062348.

[157] G. Chiribella, G. M. D'Ariano, and P. Perinotti. "Informational derivation of quantum theory". *Phys. Rev. A* 84.1 (2011). ISSN: 1094-1622. DOI: 10.1103/physreva.84.012311.

[158] L. Masanes and M. P. Müller. "A derivation of quantum theory from physical requirements". *New. J. Phys.* 13.6 (2011), p. 063001. ISSN: 1367-2630. DOI: 10.1088/1367-2630/13/6/063001.

[159] M. P. Müller and L. Masanes. "Information-Theoretic Postulates for Quantum Theory". *Quantum Theory: Informational Foundations and Foils*. Springer Netherlands, 2015, pp. 139–170. ISBN: 9789401773034. DOI: 10.1007/978-94-017-7303-4_5.

[160] J. Barrett. "Information processing in generalized probabilistic theories". *Phys. Rev. A* 75.3 (2007), p. 032304. DOI: 10.1103/PhysRevA.75.032304.

[161] B. Dakic and C. Brukner. *Quantum Theory and Beyond: Is Entanglement Special?* 2009. arXiv: 0911.0695 [quant-ph].

[162] M. Müller. "Probabilistic theories and reconstructions of quantum theory". *SciPost Phys. Lecture Notes* (2021). ISSN: 2590-1990. DOI: 10.21468/scipostphyslectnotes.28.

[163] L. Masanes, T. D. Galley, and M. P. Müller. "The measurement postulates of quantum mechanics are operationally redundant". *Nat. Commun.* 10.1 (2019). ISSN: 2041-1723. DOI: 10.1038/s41467-019-09348-x.

[164] P. Goyal, K. H. Knuth, and J. Skilling. "Origin of complex quantum amplitudes and Feynman's rules". *Phys. Rev. A* 81.2 (2010). ISSN: 1094-1622. DOI: 10.1103/physreva.81.022109.

[165] S. Popescu and D. Rohrlich. "Quantum nonlocality as an axiom". *Found. Phys.* 24.3 (1994), pp. 379–385. DOI: 10.1007/BF02058098.

[166] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger. "Limit on Nonlocality in Any World in Which Communication Complexity Is Not Trivial". *Phys. Rev. Lett.* 96.25 (2006). ISSN: 1079-7114. DOI: 10.1103/physrevlett.96.250401.

[167] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski. "Information causality as a physical principle". *Nature* 461.7267 (2009), pp. 1101–1104.

[168] T. Fritz, A. Sainz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acín. "Local orthogonality as a multipartite principle for quantum correlations". *Nat. Commun.* 4.1 (2013). ISSN: 2041-1723. DOI: 10.1038/ncomms3263.

[169] J. Oppenheim and S. Wehner. "The Uncertainty Principle Determines the Nonlocality of Quantum Mechanics". *Science* 330.6007 (2010), pp. 1072–1074. ISSN: 1095-9203. DOI: 10.1126/science.1192065.

[170] R. D. Sorkin. "Quantum mechanics as quantum measure theory". *Modern Phys. Lett. A* 09.33 (1994), pp. 3119–3127. ISSN: 1793-6632. DOI: 10.1142/s0217732394002940x.

[171] Č. Brukner and A. Zeilinger. "Information Invariance and Quantum Probabilities". *Found. Phys.* 39.7 (2009), pp. 677–689. ISSN: 1572-9516. DOI: 10.1007/s10701-009-9316-7.

[172] R. Clifton, J. Bub, and H. Halvorson. "Characterizing Quantum Theory in Terms of Information-Theoretic Constraints". *Found. Phys.* 33.11 (2003), pp. 1561–1591. ISSN: 1572-9516. DOI: 10.1023/a:1026056716397.

[173] A. Cabello. "Quantum correlations from simple assumptions". *Phys. Rev. A* 100 (3 2019), p. 032120. DOI: 10.1103/PhysRevA.100.032120.

[174] A. Cabello. "The problem of quantum correlations and the totalitarian principle". *Philos. Trans. Royal Soc. A* 377.2157 (2019), p. 20190136. ISSN: 1471-2962. DOI: 10.1098/rsta.2019.0136.

[175] G. Chiribella and X. Yuan. *Measurement sharpness cuts nonlocality and contextuality in every physical theory*. 2014. arXiv: 1404.3348 [quant-ph].

[176] P. A. Höhn. "Toolbox for reconstructing quantum theory from rules on information acquisition". *Quantum* 1 (2017), p. 38. ISSN: 2521-327X. DOI: 10.22331/q-2017-12-14-38.

[177] P. A. Höhn. "Quantum theory from rules on information acquisition". *Entropy* 19 (2017), p. 98. DOI: 10.3390/e19030098. arXiv: 1612.06849 [quant-ph].

[178] P. A. Höhn and C. S. P. Wever. "Quantum theory from questions". *Phys. Rev. A* 95.1 (2017). ISSN: 2469-9934. DOI: 10.1103/physreva.95.012102.

[179] M. P. Müller. "Law without law: from observer states to physics via algorithmic information theory". *Quantum* 4 (2020), p. 301. ISSN: 2521-327X. DOI: 10.22331/q-2020-07-20-301.

[180] C. A. Fuchs and R. Schack. "Quantum-Bayesian coherence". *Rev. Modern Phys.* 85.4 (2013), pp. 1693–1715. ISSN: 1539-0756. DOI: 10.1103/revmodphys.85.1693.

[181] J. H. Selby, C. M. Scandolo, and B. Coecke. "Reconstructing quantum theory from diagrammatic postulates". *Quantum* 5 (2021), p. 445. ISSN: 2521-327X. DOI: 10.22331/q-2021-04-28-445.

[182] B. Coecke. "Quantum picturalism". *Contemporary Physics* 51.1 (2009), pp. 59–83. ISSN: 1366-5812. DOI: 10.1080/00107510903257624.

[183] S. Abramsky and B. Coecke. "A categorical semantics of quantum protocols". *Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science, 2004.* IEEE. 2004, pp. 415–425. DOI: 10.48550/arXiv.quant-ph/0402130.

[184] S. Abramsky and B. Coecke. "Categorical quantum mechanics". *Handbook of quantum logic and quantum structures* 2 (2009), pp. 261–325. DOI: 10.48550/arXiv.0808.1023.

[185] V. Fock. "Über die invariante Form der Wellen- und der Bewegungsgleichungen für einen geladenen Massenpunkt". *Z. Physik* 39 (1926), pp. 226–232. DOI: 10.1007/BF01321989.

[186] O. Klein. "Elektrodynamik und Wellenmechanik vom Standpunkt des Korrespondenzprinzips". *Z. Physik* 41 (1927), pp. 407–442. DOI: 10.1007/BF01400205.

[187] W. Gordon. "Der Comptoneffekt nach der Schrödingerschen Theorie". *Z. Phys.* 40.1 (1926), pp. 117–133. DOI: 10.1007/BF01390840.

[188] P. A. M. Dirac. "The Quantum Theory of the Electron". *Proc. R. Soc. Lond., A* 117.778 (1928), pp. 610–624. ISSN: 09501207.

[189] E. P. Wigner. "On Unitary Representations of the Inhomogeneous Lorentz Group". *Annals Math.* 40 (1939). Ed. by Y. S. Kim and W. W. Zachary, pp. 149–204. DOI: 10.2307/1968551.

[190] K. G. Wilson. "Renormalization group and critical phenomena. 1. Renormalization group and the Kadanoff scaling picture". *Phys. Rev. B* 4 (1971), pp. 3174–3183. DOI: 10.1103/PhysRevB.4.3174.

[191] J. Polchinski. "Renormalization and effective lagrangians". *Nucl. Phys. B* 231.2 (1984), pp. 269–295. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(84)90287-6.

[192] N. Dupuis, L. Canet, A. Eichhorn, W. Metzner, J. Pawlowski, M. Tissier, and N. Wschebor. "The nonperturbative functional renormalization group and its applications". *Phys. Rep.* 910 (2021), pp. 1–114. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2021.01.001.

[193] B. S. Kay and R. M. Wald. "Theorems on the uniqueness and thermal properties of stationary, nonsingular, quasifree states on spacetimes with a bifurcate killing horizon". *Phys. Rep.* 207.2 (1991), pp. 49–136. ISSN: 0370-1573. DOI: 10.1016/0370-1573(91)90015-E.

[194] M. J. Radzikowski. "Micro-local approach to the Hadamard condition in quantum field theory on curved space-time". *Commun. Math. Phys.* 179 (1996), pp. 529–553. DOI: 10.1007/BF02100096.

[195] R. Brunetti and K. Fredenhagen. "Microlocal Analysis and Interacting Quantum Field Theories: Renormalization on Physical Backgrounds". *Commun. Math. Phys.* 208.3 (2000), pp. 623–661. ISSN: 1432-0916. DOI: 10.1007/s002200050004.

[196] C. J. Fewster and R. Verch. "The necessity of the Hadamard condition". *Class. Quantum Grav.* 30.23 (2013), p. 235027. ISSN: 1361-6382. DOI: 10.1088/0264-9381/30/23/235027.

[197] N. D. Birrell and P. C. W. Davies. *Quantum Fields in Curved Space*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1982. DOI: 10.1017/CBO9780511622632.

[198] R. Brunetti, K. Fredenhagen, and R. Verch. "The Generally Covariant Locality Principle – A New Paradigm for Local Quantum Field Theory". *Commun. Math. Phys.* 237.1 (2003), pp. 31–68. ISSN: 1432-0916. DOI: 10.1007/s00220-003-0815-7.

[199] C. J. Fewster and R. Verch. *Algebraic quantum field theory in curved spacetimes*. 2015. arXiv: 1504.00586 [math-ph].

[200] K. Rejzner. *Perturbative Algebraic Quantum Field Theory: An Introduction for Mathematicians*. Cham: Springer International Publishing, 2016, pp. 83–93. ISBN: 978-3-319-25901-7. DOI: 10.1007/978-3-319-25901-7_5.

[201] S. Hollands and R. M. Wald. "Local Wick Polynomials and Time Ordered Products of Quantum Fields in Curved Spacetime". *Commun. Math. Phys.* 223.2 (2001), pp. 289–326. ISSN: 1432-0916. DOI: 10.1007/s002200100540.

[202] R. P. Feynman and A. R. Hibbs. *Quantum Mechanics and Path Integrals*. New York: McGraw-Hill, 1965.

[203] G. W. Gibbons and S. W. Hawking. "Action Integrals and Partition Functions in Quantum Gravity". *Phys. Rev. D* 15 (1977), p. 2752. DOI: 10.1103/PhysRevD.15.2752.

[204] H. Kleinert. *Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics*. 2nd. World Scientific, 1995. DOI: 10.1142/2092. eprint: https://www.worldscientific.com/doi/pdf/10.1142/2092.

[205] M. Kontsevich. "Deformation Quantization of Poisson Manifolds". *Lett. Math. Phys.* 66.3 (2003), pp. 157–216. ISSN: 1573-0530. DOI: 10.1023/b:math.0000027508.00421.bf.

[206] F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer. "Deformation theory and quantization. I. Deformations of symplectic structures". *Ann. Phys.* 111.1 (1978), pp. 61–110. ISSN: 0003-4916. DOI: https://doi.org/10.1016/0003-4916(78)90224-5.

[207] P. A. Dirac. *Lectures on quantum mechanics*. Dover Publications, 1964.

[208] L. Faddeev and V. Popov. "Feynman diagrams for the Yang-Mills field". *Phys. Lett.B* 25.1 (1967), pp. 29–30. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(67)90067-6.

[209] M. Henneaux and C. Teitelboim. *Quantization of Gauge Systems*. Princeton University Press, 1992. ISBN: 9780691087757. DOI: 10.2307/j.ctv10crg0r.

[210] I. Batalin and G. Vilkovisky. "Gauge algebra and quantization". *Phys. Lett.B* 102.1 (1981), pp. 27–31. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(81)90205-7.

[211] R. Brunetti, K. Fredenhagen, and K. Rejzner. "Quantum gravity from the point of view of locally covariant quantum field theory". *Commun. Math. Phys.* 345.3 (2016), pp. 741–779. DOI: 10.1007/s00220-016-2676-x. arXiv: 1306.1058 [math-ph].

[212] T. Thiemann. *Introduction to Modern Canonical Quantum General Relativity*. 2001. arXiv: gr-qc/0110034 [gr-qc].

[213] A. Ashtekar and E. Bianchi. "A short review of loop quantum gravity". *Rep. Progr. Phys.* 84.4 (2021), p. 042001. ISSN: 1361-6633. DOI: 10.1088/1361-6633/abed91.

[214] S. Gogioso and F. Genovese. "Quantum Field Theory in Categorical Quantum Mechanics". *Electronic Proceedings in Theoretical Computer Science* 287 (2019), pp. 163–177. ISSN: 2075-2180. DOI: 10.4204/eptcs.287.9.

[215] R. A. Shaikh and S. Gogioso. *Categorical Semantics for Feynman Diagrams*. 2022. arXiv: 2205.00466 [quant-ph].

[216] M. Howard, J. Wallman, V. Veitch, and J. Emerson. "Contextuality supplies the 'magic'for quantum computation". *Nature* 510 (2014), pp. 351–355. DOI: 10.1038/nature13460.

[217] R. Raussendorf. "Contextuality in measurement-based quantum computation". *Phys. Rev. A* 88 (2 2013), p. 022322. DOI: 10.1103/PhysRevA.88.022322.

[218] M. Lostaglio and G. Senno. "Contextual advantage for state-dependent cloning". *Quantum* 4 (2020), p. 258. ISSN: 2521-327X. DOI: 10.22331/q-2020-04-27-258.

[219] S. Gupta, D. Saha, Z. Xu, A. Cabello, and A. S. Majumdar. "Quantum Contextuality Provides Communication Complexity Advantage". *Phys. Rev. Lett.* 130 (8 2023), p. 080802. DOI: 10.1103/PhysRevLett.130.080802.

[220] F. Shahandeh. *Quantum computational advantage implies contextuality*. 2021. arXiv: 2112.00024 [quant-ph].

[221] I. H. Deutsch. "Harnessing the Power of the Second Quantum Revolution". *PRX Quantum* 1 (2 2020), p. 020101. DOI: 10.1103/PRXQuantum.1.020101.

[222] R. Feynman. *The character of physical law*. Twelfth printing, 1985; First MIT Press Paperback Edition, March 1967. Printed in the United States of America. Library of Congress Catalog Card Number: 67-14527. The British Broadcasting Corporation, 1965. ISBN: 0-262-06016-7, 0-262-56003-8.

[223] C. J. Wood and R. W. Spekkens. "The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning". *New. J. Phys.* 17.3 (2015), p. 033002. ISSN: 1367-2630. DOI: 10.1088/1367-2630/17/3/033002.

[224] J. C. Pearl and E. G. Cavalcanti. "Classical causal models cannot faithfully explain Bell nonlocality or Kochen-Specker contextuality in arbitrary scenarios". *Quantum* 5 (2021), p. 518. DOI: 10.22331/q-2021-08-05-518.

[225] E. G. Cavalcanti and R. Lal. "On modifications of Reichenbach's principle of common cause in light of Bell's theorem". *J. Phys. A* 47.42 (2014), p. 424018.

[226] E. G. Cavalcanti. "Classical causal models for Bell and Kochen-Specker inequality violations require fine-tuning". *Phys. Rev. X* 8.2 (2018), p. 021018. DOI: 10.1103/PhysRevX.8.021018.

[227] R. W. Spekkens. "The Status of Determinism in Proofs of the Impossibility of a Noncontextual Model of Quantum Theory". *Found. Phys.* 44.11 (2014), pp. 1125–1155. ISSN: 1572-9516. DOI: 10.1007/s10701-014-9833-x.

[228] L. Catani and M. Leifer. "A mathematical framework for operational fine tunings". *Quantum* 7 (2023), p. 948. ISSN: 2521-327X. DOI: 10.22331/q-2023-03-16-948.

[229] R. W. Spekkens. *The ontological identity of empirical indiscernibles: Leibniz's methodological principle and its significance in the work of Einstein*. 2019. arXiv: 1909.04628 [physics.hist-ph].

[230] M. Leifer and M. Pusey. "Is a time symmetric interpretation of quantum theory possible without retrocausality?" *Proc. Royal Soc. A* 473.2202 (2017). arXiv:1607.07871 [physics, physics:quant-ph], p. 20160607. ISSN: 1364-5021, 1471-2946. DOI: 10.1098/rspa.2016.0607.

[231] M. S. Leifer. "Is the quantum state real? an extended review of ψ-ontology theorems". *quanta* (3 2014), pp. 67–155. DOI: 10.12743/quanta.v3i1.22.

[232] B. Amaral and M. T. Cunha. *On graph approaches to contextuality and their role in quantum theory*. Springer, 2018. DOI: 10.1007/978-3-319-93827-1.

[233] R. Kunjwal, C. Heunen, and T. Fritz. "Quantum realization of arbitrary joint measurability structures". *Phys. Rev. A* 89.5 (2014), p. 052126.

[234] R. Kunjwal. *A note on the joint measurability of POVMs and its implications for contextuality*. 2014. arXiv: 1403.0470 [quant-ph].

[235] Z.-P. Xu and A. Cabello. "Necessary and sufficient condition for contextuality from incompatibility". *Phys. Rev. A* 99 (2 2019), p. 020103. DOI: 10.1103/PhysRevA.99.020103.

[236] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. "Proposed Experiment to Test Local Hidden-Variable Theories". *Phys. Rev. Lett.* 23 (15 1969), pp. 880–884. DOI: 10.1103/PhysRevLett.23.880.

[237] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky. "Simple Test for Hidden Variables in Spin-1 Systems". *Phys. Rev. Lett.* 101 (2 2008), p. 020403. DOI: 10.1103/PhysRevLett.101.020403.

[238] S. Abramsky, R. S. Barbosa, K. Kishida, R. Lal, and S. Mansfield. "Possibilities determine the combinatorial structure of probability polytopes". *J. Math. Psychol.* 74 (2016), pp. 58–65. DOI: 10.1016/j.jmp.2016.03.006.

[239] R. Choudhary, R. S. Barbosa, and A. Cabello. "Lifting noncontextuality inequalities". *Phys. Rev. A* 109 (5 2024), p. 052216. DOI: 10.1103/PhysRevA.109.052216.

[240] R. S. Barbosa, M. Karvonen, and S. Mansfield. "Closing Bell: Boxing Black Box Simulations in the Resource Theory of Contextuality". *Samson Abramsky on Logic and Structure in Computer Science and Beyond*. Springer International Publishing, 2023, pp. 475–529. DOI: 10.1007/978-3-031-24117-8_13.

[241] S. Abramsky, S. Mansfield, and R. S. Barbosa. "The Cohomology of Non-Locality and Contextuality". *Electronic Proceedings in Theoretical Computer Science* 95 (2012), pp. 1–14. DOI: 10.4204/eptcs.95.1.

[242] R. S. Barbosa, T. Douce, P.-E. Emeriau, E. Kashefi, and S. Mansfield. "Continuous-Variable Nonlocality and Contextuality". *Commun. Math. Phys.* 391.3 (2022), pp. 1047–1089. DOI: 10.1007/s00220-021-04285-7.

[243] M. Araújo, M. T. Quintino, C. Budroni, M. T. Cunha, and A. Cabello. "All noncontextuality inequalities for the n-cycle scenario". *Phys. Rev. A* 88 (2 2013), p. 022118. DOI: 10.1103/PhysRevA.88.022118.

[244] R. Wagner, R. S. Barbosa, and E. F. Galvão. "Inequalities witnessing coherence, nonlocality, and contextuality". *Phys. Rev. A* 109 (3 2024), p. 032220. DOI: 10.1103/PhysRevA.109.032220.

[245] A. Cabello. "Experimentally Testable State-Independent Quantum Contextuality". *Phys. Rev. Lett.* 101.21 (2008). ISSN: 1079-7114. DOI: 10.1103/physrevlett.101.210401.

[246] R. Choudhary. "Classical, Quantum, and Post-Quantum Aspects of Kochen-Specker Contextuality Scenarios". PhD thesis. Universidade do Minho (Portugal), 2025.

[247] A. Fine. "Hidden Variables, Joint Probability, and the Bell Inequalities". *Phys. Rev. Lett.* 48 (5 1982), pp. 291–295. DOI: 10.1103/PhysRevLett.48.291.

[248] T. Heinosaari, D. Reitzner, and P. Stano. "Notes on Joint Measurability of Quantum Observables". *Found. Phys.* 38.12 (2008), pp. 1133–1147. ISSN: 1572-9516. DOI: 10.1007/s10701-008-9256-7.

[249] R. Kunjwal. "Beyond the Cabello-Severini-Winter framework: Making sense of contextuality without sharpness of measurements". *Quantum* 3 (2019), p. 184. DOI: 10.22331/q-2019-09-09-184.

[250] G. Chiribella and X. Yuan. "Bridging the gap between general probabilistic theories and the device-independent framework for nonlocality and contextuality". *Information and Computation* 250 (2016), pp. 15–49. ISSN: 0890-5401. DOI: 10.1016/j.ic.2016.02.006.

[251] Y.-C. Liang, R. W. Spekkens, and H. M. Wiseman. "Specker's parable of the overprotective seer: A road to contextuality, nonlocality and complementarity". *Phys. Rep.* 506.1 (2011), pp. 1–39. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2011.05.001.

[252] R. Kunjwal and R. W. Spekkens. "From the Kochen-Specker Theorem to Noncontextuality Inequalities without Assuming Determinism". *Phys. Rev. Lett.* 115.11 (2015). ISSN: 1079-7114. DOI: 10.1103/physrevlett.115.110403.

[253] R. Kunjwal and S. Ghosh. "Minimal state-dependent proof of measurement contextuality for a qubit". *Phys. Rev. A* 89 (4 2014), p. 042118. DOI: 10.1103/PhysRevA.89.042118.

[254] D. Schmid, R. W. Spekkens, and E. Wolfe. "All the noncontextuality inequalities for arbitrary prepare-and-measure experiments with respect to any fixed set of operational equivalences". *Phys. Rev. A* 97.6 (2018), p. 062103. DOI: 10.1103/PhysRevA.97.062103.

[255] D. Schmid, J. H. Selby, E. Wolfe, R. Kunjwal, and R. W. Spekkens. "Characterization of noncontextuality in the framework of generalized probabilistic theories". *PRX Quantum* 2.1 (2021), p. 010331. DOI: 10.1103/PRXQuantum.2.010331.

[256] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens. "An experimental test of noncontextuality without unphysical idealizations". *Nat. Commun.* 7.1 (2016), ncomms11780. ISSN: 2041-1723. DOI: 10.1038/ncomms11780.

[257] A. Chaturvedi, M. Farkas, and V. J. Wright. "Characterising and bounding the set of quantum behaviours in contextuality scenarios". *Quantum* 5 (2021), p. 484. DOI: 10.22331/q-2021-06-29-484.

[258] D. Schmid, R. D. Baldijão, J. H. Selby, A. B. Sainz, and R. W. Spekkens. *Noncontextuality inequalities for prepare-transform-measure scenarios*. 2024. arXiv: 2407.09624 [quant-ph].

[259] N. Harrigan and R. W. Spekkens. "Einstein, incompleteness, and the epistemic view of quantum states". *Found. Phys.* 40 (2010), pp. 125–157. DOI: 10.1007/s10701-009-9347-0.

[260] M. F. Pusey. "Robust preparation noncontextuality inequalities in the simplest scenario". *Phys. Rev. A* 98.2 (2018), p. 022112. DOI: 0.1103/PhysRevA.98.022112.

[261] M. F. Pusey, L. del Rio, and B. Meyer. *Contextuality without access to a tomographically complete set*. 2019. arXiv: 1904.08699 [quant-ph].

[262] M. Khoshbin, L. Catani, and M. Leifer. "Alternative robust ways of witnessing nonclassicality in the simplest scenario". *Phys. Rev. A* 109 (3 2024), p. 032212. DOI: 10.1103/PhysRevA.109.032212.

[263] D. Deutsch. "Quantum theory, the Church–Turing principle and the universal quantum computer". *Proc. R. Soc. Lond., A* 400.1818 (1985), pp. 97–117. DOI: 10.1098/rspa.1985.0070.

[264] N. Nurgalieva and R. Renner. "Testing quantum theory with thought experiments". *Contemp. Phys.* 61.3 (2020), pp. 193–216. DOI: 10.1080/00107514.2021.1880075.

[265] H. M. Wiseman, E. G. Cavalcanti, and E. G. Rieffel. "A "thoughtful" Local Friendliness no-go theorem: a prospective experiment with new assumptions to suit". *Quantum* 7 (2023), p. 1112. ISSN: 2521-327X. DOI: 10.22331/q-2023-09-14-1112.

[266] S. Gao. *Quantum theory is incompatible with relativity: A new proof beyond Bell's theorem and a test of unitary quantum theories*. 2019.

[267] Č. Brukner. *Qubits are not observers–a no-go theorem*. 2021. arXiv: 2107.03513 [quant-ph].

[268] J. Allam and A. Matzkin. "From observer-dependent facts to frame-dependent measurement records in Wigner friend scenarios". *Europhys. Lett.* 143.6 (2023), p. 60001. DOI: 10.1209/0295-5075/acfbf4.

[269] A. Di Biagio, P. Donà, and C. Rovelli. "The arrow of time in operational formulations of quantum theory". *Quantum* 5 (2021), p. 520. DOI: 10.22331/q-2021-08-09-520.

[270] R. Healey. "Quantum theory and the limits of objectivity". *Found. Phys.* 48 (2018), pp. 1568–1589. DOI: 10.1007/s10701-018-0216-6.

[271] L. Hardy. "Nonlocality for two particles without inequalities for almost all entangled states". *Phys. Rev. Lett.* 71.11 (1993), p. 1665. DOI: 10.1103/PhysRevLett.71.1665.

[272] S. Aaronson. *It's hard to think when someone Hadamards your brain*. 2018.

[273] A. Drezet. *About Wigner Friend's and Hardy's paradox in a Bohmian approach: a comment of 'Quantum theory cannot consistently describe the use of itself'*. 2018. arXiv: 1810.10917 [quant-ph].

[274] S. Fortin and O. Lombardi. *Wigner and his many friends: A new no-go result?* 2019. arXiv: 1904.07412 [quant-ph].

[275] S. Abramsky, R. S. Barbosa, K. Kishida, R. Lal, and S. Mansfield. "Contextuality, Cohomology and Paradox". *24th EACSL Annual Conference on Computer Science Logic (CSL 2015)*. Ed. by S. Kreutzer. Vol. 41. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2015, pp. 211–228. ISBN: 978-3-939897-90-3. DOI: 10.4230/LIPIcs.CSL.2015.211.

[276] D. M. Greenberger, M. A. Horne, and A. Zeilinger. "Going beyond Bell's theorem". *Bell's theorem, quantum theory, and conceptions of the universe*. Ed. by M. Kafatos. Vol. 37. Fundamental Theories of Physics. Dordrecht: Kluwer, 1989, pp. 69–72. DOI: 10.1007/978-94-017-0849-4_10.

[277] N. D. Mermin. "Simple unified form for the major no-hidden-variables theorems". *Phys. Rev. Lett.* 65.27 (1990), p. 3373. DOI: 10.1103/PhysRevLett.65.3373.

[278] R. S. Barbosa. "Contextuality in quantum mechanics and beyond". DPhil thesis. University of Oxford, 2015.

[279] S. Abramsky, R. S. Barbosa, and S. Mansfield. "Contextual fraction as a measure of contextuality". *Phys. Rev. Lett.* 119.5 (2017), p. 050504. DOI: 10.1103/PhysRevLett.119.050504.

[280] N. D. Mermin. "Quantum mysteries revisited". *Am. J. Phys.* 58.8 (1990), pp. 731–734. DOI: 10.1119/1.16503.

[281] B. Coecke and A. Kissinger. *Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning*. Cambridge University Press, 2017. DOI: 10.1017/9781316219317.

[282] O. Oreshkov, F. Costa, and Č. Brukner. "Quantum correlations with no causal order". *Nat. Commun.* 3.1 (2012), p. 1092. DOI: 10.1038/ncomms2076.

[283] A. Sudbery. "Single-world theory of the extended Wigner's friend experiment". *Found. Phys.* 47.5 (2017), pp. 658–669. DOI: 10.1007/s10701-017-0082-7.

[284] D. Lazarovici and M. Hubert. "How Quantum Mechanics can consistently describe the use of itself". *Scientific Reports* 9.1 (2019), p. 470. DOI: 10.1038/s41598-018-37535-1.

[285] A. Sudbery. *The hidden assumptions of Frauchiger and Renner*. 2019. arXiv: 1905.13248 [quant-ph].

[286] R. Fagin, J. Y. Halpern, Y. Moses, and M. Vardi. *Reasoning about knowledge*. MIT press, 2004.

[287] P. Blackburn, M. De Rijke, and Y. Venema. *Modal logic: graph. Darst*. Vol. 53. Cambridge University Press, 2001. DOI: 10.1017/CBO9781107050884.

[288] J. B. DeBrota, C. A. Fuchs, and R. Schack. "Respecting one's fellow: QBism's analysis of Wigner's friend". *Found. Phys.* 50 (2020), pp. 1859–1874. DOI: 10.1007/s10701-020-00369-x.

[289] D. V. Tausk. *A brief introduction to the foundations of quantum theory and an analysis of the Frauchiger-Renner paradox*. 2018. arXiv: 1812.11140 [quant-ph].

[290] J. Bub. "Understanding the Frauchiger–Renner argument". *Found. Phys.* 51.2 (2021), p. 36. DOI: 10.1007/s10701-021-00420-5.

[291] R. Muciño and E. Okon. "Wigner's convoluted friends". *Stud. Hist. Philos. Sci. Part B* 72 (2020), pp. 87–90. DOI: 10.1016/j.shpsb.2020.07.001.

[292] C. L. Jones and M. P. Müller. *On the significance of Wigner's Friend in contexts beyond quantum foundations*. 2025. arXiv: 2402.08727 [quant-ph].

[293] F. J. Boge. "Quantum information versus epistemic logic: An analysis of the Frauchiger–Renner theorem". *Found. Phys.* 49.10 (2019), pp. 1143–1165. DOI: 10.1007/s10701-019-00298-4.

[294] P. Fraser, N. Nurgalieva, and L. Rio. "Fitch's knowability axioms are incompatible with quantum theory". *J. Philos. Logic* 52 (2023). DOI: 10.1007/s10992-023-09717-4.

[295] A. Corti, V. Fano, and G. Tarozzi. "A Logico-Epistemic Investigation of Frauchiger and Renner's Paradox". *Int. J. Theor. Phys.* 62.3 (2023), p. 54. DOI: 10.1007/s10773-023-05313-z.

[296] A. Baltag and S. Smets. "Logic meets Wigner's Friend (and their Friends)". *Int. J. Theor. Phys.* 63.4 (2024), p. 97. DOI: 10.1007/s10773-024-05611-0.

[297] M. Araújo. *The flaw in Frauchiger and Renner's argument*. 2018.

[298] L. Del Rio and R. Renner. "Reply to: Quantum mechanical rules for observed observers and the consistency of quantum theory". *Nat. Commun.* 15.1 (2024), p. 3024. DOI: 10.1038/s41467-024-47172-0.

[299] A. Di Biagio and C. Rovelli. "Stable facts, relative facts". *Found. Phys.* 51 (2021), pp. 1–13. DOI: 10.1007/s10701-021-00429-w.

[300] E. G. Cavalcanti, A. Di Biagio, and C. Rovelli. *On the consistency of relative facts*. 2023. arXiv: 2305.07343 [quant-ph].

[301] R. Gambini, L. P. García-Pintos, and J. Pullin. "Single-world consistent interpretation of quantum mechanics from fundamental time and length uncertainties". *Phys. Rev. A* 100 (1 2019), p. 012113. DOI: 10.1103/PhysRevA.100.012113.

[302] M. Losada, R. Laura, and O. Lombardi. "Frauchiger-Renner argument and quantum histories". *Phys. Rev. A* 100.5 (2019), p. 052114. DOI: 10.1103/PhysRevA.100.052114.

[303] A. Relaño. *Decoherence allows quantum theory to describe the use of itself*. 2018. arXiv: 1810.07065 [quant-ph].

[304] A. Relaño. "Decoherence framework for Wigner's-friend experiments". *Phys. Rev. A* 101.3 (2020), p. 032107. DOI: 10.1103/PhysRevA.101.032107.

[305] R. Kastner. "Unitary-only quantum theory cannot consistently describe the use of itself: On the Frauchiger–Renner paradox". *Found. Phys.* 50.5 (2020), pp. 441–456. DOI: 10.1007/s10701-020-00336-6.

[306] M. Schlosshauer. "Quantum decoherence". *Phys. Rep.* 831 (2019), pp. 1–57. DOI: 10.1016/j.physrep.2019.10.001.

[307] M. Schlosshauer. "Decoherence: the Quantum-to-Classical Transition". *The Frontiers Collection (Springer-Verlag, 2007)* (2007). DOI: 10.1007/978-3-540-35775-9.

[308] G. C. Ghirardi, A. Rimini, and T. Weber. "A model for a unified quantum description of macroscopic and microscopic systems". *Quantum Probability and Applications II: Proceedings of a Workshop held in Heidelberg, West Germany, October 1–5, 1984*. Springer. 1985, pp. 223–232. DOI: 10.1007/BFb0074474.

[309] G. C. Ghirardi, A. Rimini, and T. Weber. "Unified dynamics for microscopic and macroscopic systems". *Phys. Rev. D* 34.2 (1986), pp. 470–491. DOI: 10.1103/PhysRevD.34.470.

[310] R. Penrose. "On gravity's role in quantum state reduction". *General relativity and gravitation* 28 (1996), pp. 581–600. DOI: 10.1007/BF02105068.

[311] L. Diosi. "A universal master equation for the gravitational violation of quantum mechanics". *Phys. Lett.A* 120.8 (1987), pp. 377–381. DOI: 10.1016/0375-9601(87)90681-5.

[312] A. Bassi and G. Ghirardi. "Dynamical reduction models". *Phys. Rep.* 379.5-6 (2003), pp. 257–426. DOI: 10.1016/S0370-1573%2803%2900103-0.

[313] S. Forstner, M. Zych, S. Basiri-Esfahani, K. E. Khosla, and W. P. Bowen. "Nanomechanical test of quantum linearity". *Optica* 7.10 (2020), pp. 1427–1434. DOI: 10.1364/OPTICA.391671.

[314] I. Arnquist, F. Avignone III, A. Barabash, C. Barton, K. Bhimani, E. Blalock, B. Bos, M. Busch, M. Buuck, T. Caldwell, et al. "Search for spontaneous radiation from wave function collapse in the majorana demonstrator". *Phys. Rev. Lett.* 129.8 (2022), p. 080401. DOI: 10.1103/PhysRevLett.129.080401.

[315] S. Donadi, K. Piscicchia, R. Del Grande, C. Curceanu, M. Laubenstein, and A. Bassi. "Novel CSL bounds from the noise-induced radiation emission from atoms". *The European Physical Journal C* 81.8 (2021), pp. 1–10. DOI: 10.1140/epjc/s10052-021-09556-0.

[316] M. Carlesso, S. Donadi, L. Ferialdi, M. Paternostro, H. Ulbricht, and A. Bassi. "Present status and future challenges of non-interferometric tests of collapse models". *Nat. Phys.* 18.3 (2022), pp. 243–250. DOI: 10.1038/s41567-021-01489-5.

[317] J. M. Renes. *Consistency in the description of quantum measurement: Quantum theory can consistently describe the use of itself*. 2021. arXiv: 2107.02193 [quant-ph].

[318] V. Narasimhachar. *Agents governed by quantum mechanics can use it intersubjectively and consistently*. 2020. arXiv: 2010.01167 [quant-ph].

[319] A. P. Polychronakos. "Quantum mechanical rules for observed observers and the consistency of quantum theory". *Nat. Commun.* 15.1 (2024), p. 3023. DOI: 10.1038/s41467-024-47170-2.

[320] C. Rovelli. "Relational quantum mechanics". *Int. J. Theor. Phys.* 35 (1996), pp. 1637–1678. DOI: 10.1007/BF02302261.

[321] C. Rovelli. "Space is blue and birds fly through it". *Philos. Trans. Royal Soc. A* 376.2123 (2018), p. 20170312. DOI: 10.1098/rsta.2017.0312.

[322] J. Lawrence, M. Markiewicz, and M. Żukowski. "Relative facts of relational quantum mechanics are incompatible with quantum mechanics". *Quantum* 7 (2023), p. 1015. DOI: 10.22331/q-2023-05-23-1015.

[323] E. Adlam and C. Rovelli. *Information is physical: Cross-perspective links in relational quantum mechanics*. 2022. arXiv: 2203.13342 [quant-ph].

[324] L. Hardy. "Are quantum states real?" *Int. J. Modern Phys. B* 27.01n03 (2013), p. 1345012. DOI: 10.1142/S0217979213450124.

[325] N. Ormrod and J. Barrett. *Quantum influences and event relativity*. 2024. arXiv: 2401.18005 [quant-ph].

[326] E. G. Cavalcanti. "The view from a Wigner bubble". *Found. Phys.* 51.2 (2021), p. 39. DOI: 10.1007/s10701-021-00417-0.

[327] P. Martin-Dussaud, T. Carette, J. Głowacki, V. Zatloukal, and F. Zalamea. "Fact-nets: towards a mathematical framework for relational quantum mechanics". *Found. Phys.* 53.1 (2023), p. 26. DOI: 10.1007/s10701-022-00653-y.

[328] A.-C. de la Hamette and T. D. Galley. "Quantum reference frames for general symmetry groups". *Quantum* 4 (2020), p. 367. ISSN: 2521-327X. DOI: 10.22331/q-2020-11-30-367.

[329] F. Giacomini, E. Castro-Ruiz, and Č. Brukner. "Quantum mechanics and the covariance of physical laws in quantum reference frames". *Nat. Commun.* 10.1 (2019), p. 494. DOI: 10.1038/s41467-018-08155-0.

[330] H. M. Wiseman, S. J. Jones, and A. C. Doherty. "Steering, entanglement, nonlocality, and the Einstein-Podolsky-Rosen paradox". *Phys. Rev. Lett.* 98.14 (2007), p. 140402. DOI: 10.1103/PhysRevLett.98.140402.

[331] E. Schrödinger. "Discussion of probability relations between separated systems". *Math. Proc. Camb. Philos. Soc.* Vol. 31. Cambridge University Press. 1935, pp. 555–563. DOI: 10.1017/S0305004100013554.

[332] O. Gühne, E. Haapasalo, T. Kraft, J.-P. Pellonpää, and R. Uola. "Colloquium: Incompatible measurements in quantum information science". *Rev. Modern Phys.* 95.1 (2023), p. 011003. DOI: 10.1103/RevModPhys.95.011003.

[333] A. Streltsov, G. Adesso, and M. B. Plenio. "Colloquium: Quantum coherence as a resource". *Rev. Modern Phys.* 89.4 (2017), p. 041003. DOI: 10.1103/RevModPhys.89.041003.

[334] E. F. Galvão and D. J. Brod. "Quantum and classical bounds for two-state overlaps". *Phys. Rev. A* 101.6 (2020), p. 062110. DOI: 10.1103/PhysRevA.101.062110.

[335] A. Cabello, P. Badzia ˛g, M. Terra Cunha, and M. Bourennane. "Simple Hardy-Like Proof of Quantum Contextuality". *Phys. Rev. Lett.* 111 (18 2013), p. 180404. DOI: 10.1103/PhysRevLett.111.180404.

[336] L. Santos and B. Amaral. "Conditions for logical contextuality and nonlocality". *Phys. Rev. A* 104 (2 2021), p. 022201. DOI: 10.1103/PhysRevA.104.022201.

[337] D. Schmid, Y. Yīng, and M. Leifer. *Copenhagenish interpretations of quantum mechanics*. 2025. arXiv: 2506.00112 [quant-ph].

[338] M. Pusey. *Is QBism 80% complete, or 20%?* YouTube video. 2017.

[339] M. Leifer. *What Are Copenhagenish Interpretations and Should They Be Perspectival?* YouTube video. 2020.

[340] N. Nurgalieva. "Multi-agent paradoxes in physical theories". PhD thesis. ETH Zurich, 2023.

[341] E. Woodhead. "Imperfections and self testing in prepare-and-measure quantum key distribution". PhD thesis. Laboratoire d'Information Quantique, Universite libre de Bruxelles, 2014.

[342] A. Fine. "Joint distributions, quantum correlations, and commuting observables". *J. Math. Phys.* 23.7 (1982), pp. 1306–1310. DOI: 10.1063/1.525514.

[343] N. D. Mermin. "Hidden variables and the two theorems of John Bell". *Rev. Mod. Phys.* 65 (3 1993), pp. 803–815. DOI: 10.1103/RevModPhys.65.803.

[344] A. Peres. *Quantum theory: concepts and methods*. Vol. 72. Springer, 1997.

[345] L. Catani, M. Leifer, G. Scala, D. Schmid, and R. W. Spekkens. "What is Nonclassical about Uncertainty Relations?" *Phys. Rev. Lett.* 129 (24 2022), p. 240401. DOI: 10.1103/PhysRevLett.129.240401.

[346] A. Cabello. "Converting Contextuality into Nonlocality". *Phys. Rev. Lett.* 127 (7 2021), p. 070401. DOI: 10.1103/PhysRevLett.127.070401.

[347] A. Cabello. "Proposal for Revealing Quantum Nonlocality via Local Contextuality". *Phys. Rev. Lett.* 104 (22 2010), p. 220401. DOI: 10.1103/PhysRevLett.104.220401.

[348] L. Aolita, R. Gallego, A. Acín, A. Chiuri, G. Vallone, P. Mataloni, and A. Cabello. "Fully nonlocal quantum correlations". *Phys. Rev. A* 85 (3 2012), p. 032107. DOI: 10.1103/PhysRevA.85.032107.

[349] A. Cabello and M. T. Cunha. "Proposal of a Two-Qutrit Contextuality Test Free of the Finite Precision and Compatibility Loopholes". *Phys. Rev. Lett.* 106 (19 2011), p. 190401. DOI: 10.1103/PhysRevLett.106.190401.

[350] L. Henaut, L. Catani, D. E. Browne, S. Mansfield, and A. Pappa. "Tsirelson's bound and Landauer's principle in a single-system game". *Phys. Rev. A* 98 (6 2018), p. 060302. DOI: 10.1103/PhysRevA.98.060302.

[351] L. Catani, R. Faleiro, P.-E. Emeriau, S. Mansfield, and A. Pappa. "Connecting XOR and XOR* games". *Phys. Rev. A* 109 (1 2024), p. 012427. DOI: 10.1103/PhysRevA.109.012427.

[352] V. J. Wright and M. Farkas. "Invertible Map between Bell Nonlocal and Contextuality Scenarios". *Phys. Rev. Lett.* 131 (22 2023), p. 220202. DOI: 10.1103/PhysRevLett.131.220202.

[353] V. Gitton and M. P. Woods. "System loophole of generalized noncontextuality". *Phys. Rev. Res.* 6.4 (2024). ISSN: 2643-1564. DOI: 10.1103/physrevresearch.6.043289.

[354] H. M. Wiseman and E. G. Cavalcanti. "Causarum Investigatio and the two Bell's theorems of John Bell". *Quantum [Un] Speakables II: Half a Century of Bell's Theorem* (2017), pp. 119–142. DOI: https://doi.org/10.48550/arXiv.1503.06413.

[355] D. Schmid, J. H. Selby, and R. W. Spekkens. *Unscrambling the omelette of causation and inference: The framework of causal-inferential theories*. 2021. arXiv: 2009.03297 [quant-ph].

[356] A. Chaturvedi and D. Saha. "Quantum prescriptions are more ontologically distinct than they are operationally distinguishable". *Quantum* 4 (2020), p. 345. ISSN: 2521-327X. DOI: 10.22331/q-2020-10-21-345.

[357] N. Nurgalieva and V. Vilasini. *Any theory that admits a Wigner's Friend type multi-agent paradox is logically contextual*. 2025. arXiv: 2502.03874 [quant-ph].

[358] S. Mansfield. "Reality of the quantum state: Towards a stronger ψ-ontology theorem". *Phys. Rev. A* 94.4 (2016), p. 042124. ISSN: 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.94.042124.

[359] R. Spekkens. *Why I am not a psi-ontologist*. 2012. URL: https://pirsa.org/12050021/.

[360] A. Elga. "Self-locating belief and the Sleeping Beauty problem". *Analysis* 60.2 (2000), pp. 143–147. DOI: 10.1093/analys/60.2.143.

[361] R. J. Aumann, S. Hart, and M. Perry. "The forgetful passenger". *Games and Economic Behavior* 20.1 (1997), pp. 117–120. ISSN: 0899-8256. DOI: 10.1006/game.1997.0578.

[362] M. Piccione and A. Rubinstein. "On the interpretation of decision problems with imperfect recall". *Games and Economic Behavior* 20.1 (1997), pp. 3–24. ISSN: 0899-8256. DOI: 10.1006/game.1997.0536.

[363] D. Lewis. "Sleeping Beauty: reply to Elga". *Analysis* 61.3 (2001), pp. 171–176. DOI: 10.1093/analys/61.3.171.

[364] A. Elga. "Defeating Dr. Evil with self-locating belief". *Philosophy and Phenomenological Research* 69.2 (2004), pp. 383–396. DOI: 10.1111/j.1933-1592.2004.tb00400.x.

[365] C. Hitchcock. "Beauty and the bets". *Synthese* 139 (2004), pp. 405–420. DOI: 10.1023/B:SYNT.0000024889.29125.c0.

[366] K. Draper and J. Pust. "Diachronic Dutch books and Sleeping Beauty". *Synthese* 164 (2008), pp. 281–287. DOI: 10.1007/s11229-007-9226-1.

[367] F. Arntzenius. "Some problems for conditionalization and reflection". *J. Philos.* 100.7 (2003), pp. 356–370. DOI: 10.1007/978-3-319-20451-2_10.

[368] C. Dorr. "Sleeping Beauty: in defence of Elga". *Analysis* 62.4 (2002), pp. 292–296. DOI: 10.1093/analys/62.4.292.

[369] S. van der Meijden. "On the Sleeping Beauty problem". B.S. thesis. Utrecht University, 2013.

[370] T. Horgan. "Sleeping Beauty awakened: New odds at the dawn of the new day". *Analysis* 64.1 (2004), pp. 10–21. DOI: 10.1093/analys/64.1.10.

[371] R. Weintraub. "Sleeping Beauty: a simple solution". *Analysis* 64.1 (2004), pp. 8–10. DOI: 10.1093/analys/64.1.8.

[372] B. Monton. "Sleeping Beauty and the forgetful Bayesian". *Analysis* 62.1 (2002), pp. 47–53. DOI: 10.1093/analys/62.1.47.

[373] M. G. Titelbaum. "An embarrassment for double-halfers". *Thought: J. Philos.* 1.2 (2012), pp. 146–151. DOI: 10.1002/tht3.21.

[374] M. G. Titelbaum. *Quitting certainties: A Bayesian framework modeling degrees of belief*. 2013. DOI: 10.1017/S0266267114000492.

[375] R. M. Neal. *Puzzles of anthropic reasoning resolved using full non-indexical conditioning*. 2006. arXiv: math/0608592 [math.ST].

[376] R. Stalnaker. *Our knowledge of the internal world*. 2008. DOI: 10.1093/acprof:oso/9780199545995.001.0001.

[377] B. Groisman. "The end of Sleeping Beauty's nightmare". *The British Journal for the Philosophy of Science* 59.3 (2008), pp. 409–416. DOI: 10.1093/bjps/axn015.

[378] D. Bradley and H. Leitgeb. "When betting odds and credences come apart: More worries for Dutch book arguments". *Analysis* 66.2 (2006), pp. 119–127. ISSN: 0003-2638. DOI: 10.1093/analys/66.2.119.

[379] J. Rosenthal. "A Mathematical Analysis of the Sleeping Beauty Problem". *The Mathematical Intelligencer* 31 (2009), pp. 32–37. DOI: 10.1007/s00283-009-9060-z.

[380] T. Horgan. "The Exchange Continued: Pust's Response to My Reply". *Essays on Paradoxes*. Ed. by T. Horgan. Oxford University Press, 2017, pp. 226–246.

[381] M. G. Titelbaum. "Ten reasons to care about the Sleeping Beauty problem". *Philosophy Compass* 8.11 (2013), pp. 1003–1017. DOI: 10.1111/phc3.12080.

[382] S. M. Carroll. *Why Boltzmann brains are bad*. 2020. DOI: 10.48550/arXiv.1702.00850.

[383] R. White. "The generalized Sleeping Beauty problem: a challenge for thirders". *Analysis* 66.2 (2006), pp. 114–119. DOI: 10.1093/analys/66.2.114.

[384] T. Horgan. "Synchronic Bayesian Updating and the Generalized Sleeping Beauty Problem". *Analysis* 67.1 (2007), pp. 50–59. ISSN: 00032638, 14678284. DOI: 10.1093/analys/67.1.50.

[385] C. J. Meacham. "Sleeping Beauty and the dynamics of de se beliefs". *Philosophical Studies* 138.2 (2008), pp. 245–269. DOI: 10.1007/s11098-006-9036-1.

[386] L. Walleghem. *Wigner's friend's black hole adventure: an argument for complementarity?* 2025. arXiv: 2507.05369 [gr-qc].

[387] L. Walleghem and C. Cepollaro. *Particle detector in a position-superposed black hole spacetime*. 2026. arXiv: 2604.11897 [quant-ph].

[388] J. Polchinski. "The Black Hole Information Problem". *New Frontiers in Fields and Strings*. World Scientific, 2016. DOI: 10.1142/9789813149441_0006.

[389] D. N. Page. "Average entropy of a subsystem". *Phys. Rev. Lett.* 71.9 (1993), p. 1291. DOI: 10.1103/PhysRevLett.71.1291.

[390] C. Holzhey, F. Larsen, and F. Wilczek. "Geometric and renormalized entropy in conformal field theory". *Nucl. Phys. B* 424.3 (1994), pp. 443–467. ISSN: 05503213. DOI: 10.1016/0550-3213(94)90402-2.

[391] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini. "The entropy of Hawking radiation". *Rev. Modern Phys.* 93.3 (2021), p. 035002. DOI: 10.1103/RevModPhys.93.035002.

[392] J. Eisert, M. Cramer, and M. B. Plenio. "Colloquium: Area laws for the entanglement entropy". *Rev. Mod. Phys.* 82 (1 2010), pp. 277–306. DOI: 10.1103/RevModPhys.82.277.

[393] S. D. Mathur. "The information paradox: a pedagogical introduction". *Class. Quantum Gravity* 26.22 (2009), p. 224001. DOI: 10.1088/0264-9381/26/22/224001.

[394] C. R. Stephens, G. 't Hooft, and B. F. Whiting. "Black hole evaporation without information loss". *Class. Quantum Grav.* 11.3 (1994), pp. 621–647. ISSN: 1361-6382. DOI: 10.1088/0264-9381/11/3/014.

[395] T. G. Mertens and G. J. Turiaci. "Solvable models of quantum black holes: a review on Jackiw–Teitelboim gravity". *Living Rev. Relativ.* 26.1 (2023), p. 4. DOI: 10.1007/s41114-023-00046-1.

[396] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini. "Replica wormholes and the entropy of Hawking radiation". *J. High Energy Phys.* 2020.5 (2020), pp. 1–42. DOI: 10.1007/jhep05(2020)013.

[397] G. Penington, S. H. Shenker, D. Stanford, and Z. Yang. "Replica wormholes and the black hole interior". *J. High Energy Phys.* 2022.3 (2022), pp. 1–87. DOI: 10.48550/arXiv.1911.11977.

[398] K. Goto, T. Hartman, and A. Tajdini. "Replica wormholes for an evaporating 2D black hole". *J. High Energy Phys.* 2021.4 (2021). ISSN: 1029-8479. DOI: 10.1007/jhep04(2021)289.

[399] X. Calmet and S. D. H. Hsu. *Black Hole Information, Replica Wormholes, and Macroscopic Entanglement*. 2024. arXiv: 2412.07807 [hep-th].

[400] Y. An and P. Cheng. "Replica wormhole as a vacuum-to-vacuum transition". *The European Physical Journal C* 83.4 (2023). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-023-11518-7.

[401] A. Ashtekar, V. Taveras, and M. Varadarajan. "Information is Not Lost in the Evaporation of 2-dimensional Black Holes". *Phys. Rev. Lett.* 100.21 (2008). arXiv:0801.1811 [gr-qc], p. 211302. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.100.211302.

[402] A. Ashtekar, F. Pretorius, and F. M. Ramazanoğlu. "Surprises in the Evaporation of 2D Black Holes". *Phys. Rev. Lett.* 106 (16 2011), p. 161303. DOI: 10.1103/PhysRevLett.106.161303.

[403] A. Ashtekar, F. Pretorius, and F. M. Ramazanoğlu. "Evaporation of two-dimensional black holes". *Phys. Rev. D* 83 (4 2011), p. 044040. DOI: 10.1103/PhysRevD.83.044040.

[404] J. Barenboim, A. V. Frolov, and G. Kunstatter. "No drama in two-dimensional black hole evaporation". *Phys. Rev. Res.* 6 (3 2024), p. L032055. DOI: 10.1103/PhysRevResearch.6.L032055.

[405] P. Chen, Y. Ong, and D. Yeom. "Black hole remnants and the information loss paradox". *Phys. Rep.* 603 (2015), pp. 1–45. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2015.10.007.

[406] Y. C. Ong. *The Case For Black Hole Remnants: A Review*. 2025. arXiv: 2412.00322 [gr-qc].

[407] F. D'Ambrosio, M. Christodoulou, P. Martin-Dussaud, C. Rovelli, and F. Soltani. "End of a black hole's evaporation". *Phys. Rev. D* 103.10 (2021), p. 106014. DOI: 10.1103/PhysRevD.103.106014.

[408] M. Christodoulou, C. Rovelli, S. Speziale, and I. Vilensky. "Planck star tunneling time: An astrophysically relevant observable from background-free quantum gravity". *Phys. Rev. D* 94.8 (2016), p. 084035. DOI: 10.1103/PhysRevD.94.084035.

[409] C. Rovelli and F. Vidotto. "Planck stars". *Int. J. Modern Phys. D* 23 (2014), p. 1442026. DOI: 10.1142/S0218271814420267.

[410] E. Bianchi, M. Christodoulou, F. D'Ambrosio, H. M. Haggard, and C. Rovelli. "White Holes as Remnants: A Surprising Scenario for the End of a Black Hole". *Class. Quant. Grav.* 35.22 (2018), p. 225003. DOI: 10.1088/1361-6382/aae550.

[411] J. Ben Achour, S. Brahma, S. Mukohyama, and J. .-. Uzan. "Towards consistent black-to-white hole bounces from matter collapse". *J. Cosmol. Astropart. Phys.* 09 (2020), p. 020. DOI: 10.1088/1475-7516/2020/09/020.

[412] C. Rovelli and F. Vidotto. *Planck stars, White Holes, Remnants and Planck-mass quasi-particles. The quantum gravity phase in black holes' evolution and its manifestations*. 2024. arXiv: 2407.09584 [gr-qc].

[413] H. M. Haggard and C. Rovelli. "Quantum-gravity effects outside the horizon spark black to white hole tunneling". *Phys. Rev. D* 92.10 (2015), p. 104020. DOI: 10.1103/PhysRevD.92.104020. arXiv: 1407.0989 [gr-qc].

[414] M. Han, C. Rovelli, and F. Soltani. "Geometry of the black-to-white hole transition within a single asymptotic region". *Phys. Rev. D* 107.6 (2023), p. 064011. DOI: 10.1103/PhysRevD.107.064011. arXiv: 2302.03872 [gr-qc].

[415] M. Christodoulou and C. Rovelli. "How big is a black hole?" *Phys. Rev. D* 91.6 (2015), p. 064046. DOI: 10.1103/PhysRevD.91.064046.

[416] M. Christodoulou and T. De Lorenzo. "Volume inside old black holes". *Phys. Rev. D* 94.10 (2016), p. 104002. DOI: 10.1103/PhysRevD.94.104002.

[417] S. B. Giddings. "Why aren't black holes infinitely produced?" *Phys. Rev. D* 51.12 (1995), pp. 6860–6869. ISSN: 0556-2821. DOI: 10.1103/physrevd.51.6860.

[418] L. Susskind. *Trouble For Remnants*. 1995. arXiv: hep-th/9501106 [hep-th].

[419] S. B. Giddings. "Constraints on black hole remnants". *Phys. Rev. D* 49 (1994), pp. 947–957. DOI: 10.1103/PhysRevD.49.947. arXiv: hep-th/9304027.

[420] R. Parentani and T. Piran. "Internal Geometry of an Evaporating Black Hole". *Phys. Rev. Lett.* 73.21 (1994), pp. 2805–2808. ISSN: 0031-9007. DOI: 10.1103/physrevlett.73.2805.

[421] D. A. Easson and R. H. Brandenberger. "Universe generation from black hole interiors". *J. High Energy Phys.* 2001.06 (2001), pp. 024–024. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2001/06/024.

[422] S. Mathur. "The fuzzball proposal for black holes: an elementary review". *Fortschr. Phys.* 53.7–8 (2005), pp. 793–827. ISSN: 1521-3978. DOI: 10.1002/prop.200410203.

[423] S. B. Giddings. "Nonviolent information transfer from black holes: A field theory parametrization". *Phys. Rev. D* 88.2 (2013), p. 024018. DOI: 10.1103/PhysRevD.88.024018.

[424] P. O. Mazur and E. Mottola. "Gravitational vacuum condensate stars". *Proc. Natl. Acad. Sci.* 101.26 (2004), pp. 9545–9550. DOI: 10.1073/pnas.0402717101.

[425] H. Kawai, Y. Matsuo, and Y. Yokokura. "A self-consistent model of the black hole evaporation". *Int. J. Modern Phys. A* 28.14 (2013), p. 1350050. ISSN: 1793-656X. DOI: 10.1142/s0217751x13500504.

[426] V. Baccetti, R. B. Mann, and D. R. Terno. "Do event horizons exist?" *Int. J. Modern Phys. D* 26.12 (2017), p. 1743008. ISSN: 1793-6594. DOI: 10.1142/s0218271817430088.

[427] V. Baccetti, R. B. Mann, and D. R. Terno. "Role of evaporation in gravitational collapse". *Class. Quantum Gravity* 35.18 (2018), p. 185005. ISSN: 1361-6382. DOI: 10.1088/1361-6382/aad70e.

[428] R. Carballo-Rubio. "Stellar Equilibrium in Semiclassical Gravity". *Phys. Rev. Lett.* 120 (2018), p. 061102. ISSN: 1079-7114. DOI: 10.1103/physrevlett.120.061102.

[429] V. Cardoso and P. Pani. "Testing the nature of dark compact objects: a status report". *Living Rev. Relativ.* 22 (2019), pp. 1–104. DOI: 10.1007/s41114-019-0020-4.

[430] P. Chen, W. G. Unruh, C. Wu, and D. Yeom. "Pre-Hawking radiation cannot prevent the formation of apparent horizon". *Phys. Rev. D* 97 (2018), p. 064045. ISSN: 2470-0029. DOI: 10.1103/physrevd.97.064045.

[431] X. Calmet and S. D. Hsu. "Quantum hair and black hole information". *Phys. Lett.B* 827 (2022), p. 136995. DOI: 10.1016/j.physletb.2022.136995.

[432] S. Weinberg. "Infrared Photons and Gravitons". *Phys. Rev.* 140.2B (1965), B516–B524. ISSN: 0031-899X. DOI: 10.1103/PhysRev.140.B516.

[433] Y. B. Zel'dovich and A. G. Polnarev. "Radiation of gravitational waves by a cluster of superdense stars". *Sov. Astron.* 18 (1974), p. 17.

[434] D. Christodoulou. "Nonlinear nature of gravitation and gravitational-wave experiments". *Phys. Rev. Lett.* 67.12 (1991), p. 1486. DOI: 10.1103/PhysRevLett.67.1486.

[435] K. S. Thorne. "Gravitational-wave bursts with memory: The Christodoulou effect". *Phys. Rev. D* 45 (2 1992), pp. 520–524. DOI: 10.1103/PhysRevD.45.520.

[436] L. Bieri and A. Polnarev. "Gravitational wave displacement and velocity memory effects". *Class. Quantum Gravity* 41.13 (2024), p. 135012. ISSN: 1361-6382. DOI: 10.1088/1361-6382/ad4dfe.

[437] D. Maibach and J. Zosso. *Balance flux laws beyond general relativity*. 2026. arXiv: 2601.07091 [gr-qc].

[438] H. Bondi, M. G. J. van der Burg, and A. Metzner. "Gravitational waves in general relativity, VII. Waves from axi-symmetric isolated system". *Proc. R. Soc. London, Ser. A* 269.1336 (1962), pp. 21–52. DOI: 10.1098/rspa.1962.0161.

[439] R. K. Sachs. "Gravitational waves in general relativity VIII. Waves in asymptotically flat space-time". *Proc. R. Soc. London, Ser. A* 270.1340 (1962), pp. 103–126. DOI: 10.1098/rspa.1962.0206.

[440] G. Compère and A. Fiorucci. *Advanced Lectures on General Relativity*. 2019. arXiv: 1801.07064 [hep-th].

[441] G. Compère, S. E. Gralla, and H. Wei. "An asymptotic framework for gravitational scattering". *Class. Quantum Gravity* 40.20 (2023), p. 205018. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/acf5c1.

[442] G. Satishchandran and R. M. Wald. "Asymptotic behavior of massless fields and the memory effect". *Phys. Rev. D* 99 (8 2019), p. 084007. DOI: 10.1103/PhysRevD.99.084007.

[443] K. Prabhu and G. Satishchandran. "Infrared finite scattering theory: Amplitudes and soft theorems". *Phys. Rev. D* 110.8 (2024), p. 085022. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.110.085022.

[444] K. Prabhu, G. Satishchandran, and R. M. Wald. "Infrared finite scattering theory in quantum field theory and quantum gravity". *Phys. Rev. D* 106.6 (2022), p. 066005. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.106.066005.

[445] A. Sommerfeld. "Über die Beugung und Bremsung der Elektronen". *Ann. Phys.* 403.3 (1931), pp. 257–330. DOI: 10.1002/andp.19314030302.

[446] M. Fierz. "Über die künstliche Umwandlung des Protons in ein Neutron". PhD thesis. E. Birkhäuser & Cie., AG, 1936.

[447] F. Bloch and A. Nordsieck. "Note on the Radiation Field of the Electron". *Phys. Rev.* 52 (2 1937), pp. 54–59. DOI: 10.1103/PhysRev.52.54.

[448] N. F. Mott. "On the influence of radiative forces on the scattering of electrons". *Math. Proc. Cambridge Philos. Soc.* 27.2 (1931), pp. 255–267. DOI: 10.1017/S0305004100010379.

[449] X. Calmet, R. Casadio, S. D. Hsu, and F. Kuipers. "Quantum hair from gravity". *Phys. Rev. Lett.* 128.11 (2022), p. 111301. DOI: 10.1103/PhysRevLett.128.111301.

[450] X. Calmet, S. D. Hsu, and M. Sebastianutti. "Quantum gravitational corrections to particle creation by black holes". *Phys. Lett. B* (2023), p. 137820. DOI: 10.1016/j.physletb.2023.137820.

[451] G. Dvali and C. Gomez. *Black Hole's Quantum N-Portrait*. 2011. arXiv: 1112.3359 [hep-th].

[452] R. Eyheralde, M. Campiglia, R. Gambini, and J. Pullin. "Quantum fluctuating geometries and the information paradox". *Class. Quantum Grav.* 34.23 (2017), p. 235015. ISSN: 1361-6382. DOI: 10.1088/1361-6382/aa8e30.

[453] R. Eyheralde, R. Gambini, and J. Pullin. "Quantum fluctuating geometries and the information paradox II". *Class. Quantum Grav.* 37.6 (2020), p. 065001. ISSN: 1361-6382. DOI: 10.1088/1361-6382/ab6e89.

[454] R. Brustein and A. Medved. "Restoring predictability in semiclassical gravitational collapse". *J. High Energy Phys.* 2013.9 (2013), pp. 1–21. DOI: 10.1007/jhep09(2013)015.

[455] L. Alberte, R. Brustein, A. Khmelnitsky, and A. Medved. "Density matrix of black hole radiation". *J. High Energy Phys.* 2015.8 (2015), pp. 1–37. DOI: 10.1007/jhep08(2015)015.

[456] R. Eyheralde. "Hawking radiation from an evaporating black hole via Bogoliubov transformations". *Class. Quantum Grav.* 39.22 (2022), p. 225002. ISSN: 1361-6382. DOI: 10.1088/1361-6382/ac93e2.

[457] I. Agulló, J. Navarro-Salas, G. J. Olmo, and L. Parker. "Short-distance contribution to the spectrum of Hawking radiation". *Phys. Rev. D* 76 (4 2007), p. 044018. DOI: 10.1103/PhysRevD.76.044018.

[458] L. Donnay, G. Giribet, H. A. González, and M. Pino. "Supertranslations and Superrotations at the Black Hole Horizon". *Phys. Rev. Lett.* 116.9 (2016), p. 091101. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.116.091101.

[459] G. Barnich and C. Troessaert. "Symmetries of Asymptotically Flat Four-Dimensional Spacetimes at Null Infinity Revisited". *Phys. Rev. Lett.* 105.11 (2010), p. 111103. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.105.111103.

[460] M. Campiglia and A. Laddha. "Asymptotic symmetries and subleading soft graviton theorem". *Phys. Rev. D* 90.12 (2014), p. 124028. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.90.124028.

[461] L. Heisenberg, G. Xu, and J. Zosso. *Unifying Ordinary and Null Memory*. 2025. arXiv: 2401.05936 [gr-qc].

[462] S. Choi, A. Laddha, and A. Puhm. *The Classical Super-Rotation Infrared Triangle*. 2025. arXiv: 2412.16142.

[463] D. Yennie, S. Frautschi, and H. Suura. "The infrared divergence phenomena and high-energy processes". *Ann. Phys.* 13.3 (1961), pp. 379–452. ISSN: 0003-4916. DOI: 10.1016/0003-4916(61)90151-8.

[464] A. Vanrietvelde, P. A. Höhn, F. Giacomini, and E. Castro-Ruiz. "A change of perspective: switching quantum reference frames via a perspective-neutral framework". *Quantum* 4 (2020), p. 225. ISSN: 2521-327X. DOI: 10.22331/q-2020-01-27-225.

[465] A. Vanrietvelde, P. A. Höhn, and F. Giacomini. "Switching quantum reference frames in the N-body problem and the absence of global relational perspectives". *Quantum* 7 (2023), p. 1088. ISSN: 2521-327X. DOI: 10.22331/q-2023-08-22-1088.

[466] P. A. Höhn, A. R. H. Smith, and M. P. E. Lock. "The Trinity of Relational Quantum Dynamics". *Phys. Rev.D* 104.6 (2021), p. 066001. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.104.066001.

[467] E. Castro-Ruiz, T. D. Galley, and L. Loveridge. *Interpreting quantum reference frame transformations through a simple example*. 2025. arXiv: 2508.09540.

[468] P. A. Höhn, M. Krumm, and M. P. Müller. "Internal quantum reference frames for finite Abelian groups". *J. Math. Phys.* 63 (2022), p. 11. ISSN: 1089-7658. DOI: 10.1063/5.0088485.

[469] G. Araujo-Regado, P. A. Höhn, F. Sartini, and B. Tomova. "Soft edges: the many links between soft and edge modes". *J. High Energy Phys.* 2025.7 (2025), p. 180. DOI: 10.1007/jhep07(2025)180.

[470] V. Kabel, Č. Brukner, and W. Wieland. "Quantum reference frames at the boundary of spacetime". *Phys. Rev. D* 108.10 (2023), p. 106022. DOI: 10.1103/PhysRevD.108.106022.

[471] V. Kabel, A. de la Hamette, L. Apadula, C. Cepollaro, H. Gomes, J. Butterfield, and Č. Brukner. "Quantum coordinates, localisation of events, and the quantum hole argument". *Commun. Phys.* 8 (2025), p. 185. ISSN: 2399-3650. DOI: 10.1038/s42005-025-02084-3.

[472] C. Cepollaro and F. Giacomini. "Quantum generalisation of Einstein's equivalence principle can be verified with entangled clocks as quantum reference frames". *Class. Quantum Grav.* 41.18 (2024), p. 185009. DOI: 10.1088/1361-6382/ad6d26.

[473] D. A. Lowe, J. Polchinski, L. Susskind, L. Thorlacius, and J. Uglum. "Black hole complementarity versus locality". *Phys. Rev. D* 52 (1995), pp. 6997–7010. ISSN: 0556-2821. DOI: 10.1103/physrevd.52.6997.

[474] L. Susskind. *Black Hole Complementarity and the Harlow-Hayden Conjecture*. 2013. arXiv: 1301.4505 [hep-th].

[475] D. Harlow and P. Hayden. "Quantum Computation vs. Firewalls". *J. High Energy Phys.* 2013.6 (2013). arXiv:1301.4504 [hep-th], p. 85. ISSN: 1029-8479. DOI: 10.1007/JHEP06(2013)085.

[476] L. Susskind. "Computational complexity and black hole horizons". *Fortschritte der Physik* 64.1 (2016), pp. 24–43. DOI: 10.1002/prop.201500092.

[477] I. H. Kim, E. Tang, and J. Preskill. "The ghost in the radiation: Robust encodings of the black hole interior". *J. High Energy Phys.* 2020.6 (2020). arXiv:2003.05451 [hep-th], p. 31. ISSN: 1029-8479. DOI: 10.1007/JHEP06(2020)031.

[478] B. Yoshida. "Firewalls vs. Scrambling". *J. High Energy Phys.* 2019.10 (2019). arXiv:1902.09763 [hep-th], p. 132. ISSN: 1029-8479. DOI: 10.1007/JHEP10(2019)132.

[479] C. Akers, N. Engelhardt, D. Harlow, G. Penington, and S. Vardhan. *The black hole interior from non-isometric codes and complexity*. 2022. arXiv: 2207.06536 [hep-th].

[480] J. Oppenheim and B. Unruh. "Firewalls and flat mirrors: An alternative to the AMPS experiment which evades the Harlow-Hayden obstacle". *J. High Energy Phys.* 2014.3 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep03(2014)120.

[481] K. Papadodimas and S. Raju. "State-dependent bulk-boundary maps and black hole complementarity". *Phys. Rev. D* 89.8 (2014), p. 086010. DOI: 10.1103/PhysRevD.89.086010.

[482] K. Papadodimas and S. Raju. "Black Hole Interior in the Holographic Correspondence and the Information Paradox". *Phys. Rev. Lett.* 112 (5 2014), p. 051301. DOI: 10.1103/PhysRevLett.112.051301.

[483] D. Harlow. "Aspects of the Papadodimas-Raju proposal for the black hole interior". *J. High Energy Phys.* 2014.11 (2014). ISSN: 1029-8479. DOI: 10.1007/jhep11(2014)055.

[484] J. Cotler, P. Hayden, G. Penington, G. Salton, B. Swingle, and M. Walter. "Entanglement Wedge Reconstruction via Universal Recovery Channels". *Phys. Rev. X* 9 (2019), p. 031011. ISSN: 2160-3308. DOI: 10.1103/physrevx.9.031011.

[485] S. Lloyd and J. Preskill. "Unitarity of black hole evaporation in final-state projection models". *J. High Energy Phys.* 2014 (2014), p. 126. ISSN: 1029-8479. DOI: 10.1007/jhep08(2014)126.

[486] R. Bousso and D. Stanford. "Measurements without probabilities in the final state proposal". *Phys. Rev. D* 89 (2014), p. 044038. ISSN: 1550-2368. DOI: 10.1103/physrevd.89.044038.

[487] S. B. Giddings. "Nonviolent nonlocality". *Phys. Rev. D* 88 (2013), p. 064023. ISSN: 1550-2368. DOI: 10.1103/physrevd.88.064023.

[488] B. Concepcion, Y. Nomura, K. Ritchie, and S. Weiss. "Complementarity for a dynamical black hole". *Phys. Rev. D* 110 (8 2024), p. 086002. DOI: 10.1103/PhysRevD.110.086002.

[489] N. Zilberman, M. Casals, A. Ori, and A. C. Ottewill. "Quantum Fluxes at the Inner Horizon of a Spinning Black Hole". *Phys. Rev. Lett.* 129 (26 2022), p. 261102. DOI: 10.1103/PhysRevLett.129.261102.

[490] B. A. Juárez-Aubry and J. Louko. "Quantum kicks near a Cauchy horizon". *AVS Quantum Science* 4.1 (2022), p. 013201. ISSN: 2639-0213. DOI: 10.1116/5.0073373.

[491] B. A. Juárez-Aubry. "Can a particle detector cross a Cauchy horizon?" *Int. J. Modern Phys. D* 24.09 (2015), p. 1542005. ISSN: 1793-6594. DOI: 10.1142/s0218271815420055.

[492] B. A. Juárez-Aubry. "Quantum strong cosmic censorship and black hole evaporation". *Class. Quantum Grav.* 41.19 (2024), p. 195027. ISSN: 1361-6382. DOI: 10.1088/1361-6382/ad756c.

[493] A. Lanir, A. Ori, N. Zilberman, O. Sela, A. Maline, and A. Levi. "Analysis of quantum effects inside spherical charged black holes". *Phys. Rev. D* 99 (6 2019), p. 061502. DOI: 10.1103/PhysRevD.99.061502.

[494] S. Hollands, C. Klein, and J. Zahn. "Quantum stress tensor at the Cauchy horizon of the Reissner–Nordström–de Sitter spacetime". *Phys. Rev. D* 102 (8 2020), p. 085004. DOI: 10.1103/PhysRevD.102.085004.

[495] S. Hollands, R. M. Wald, and J. Zahn. "Quantum instability of the Cauchy horizon in Reissner–Nordström–deSitter spacetime". *Class. Quantum Grav.* 37.11 (2020), p. 115009. ISSN: 1361-6382. DOI: 10.1088/1361-6382/ab8052.

[496] K. Papadodimas, S. Raju, and P. Shrivastava. "A simple quantum test for smooth horizons". *J. High Energy Phys.* 2020.12 (2020). ISSN: 1029-8479. DOI: 10.1007/jhep12(2020)003.

[497] V. Cardoso, J. L. Costa, K. Destounis, P. Hintz, and A. Jansen. "Quasinormal Modes and Strong Cosmic Censorship". *Phys. Rev. Lett.* 120 (3 2018), p. 031103. DOI: 10.1103/PhysRevLett.120.031103.

[498] P. Shrivastava. *Quantum aspects of charged black holes in de Sitter space*. 2020. arXiv: 2009.03261 [hep-th].

[499] O. J. Dias, H. S. Reall, and J. E. Santos. "The BTZ black hole violates strong cosmic censorship". *J. High Energy Phys.* 2019.12 (2019). ISSN: 1029-8479. DOI: 10.1007/jhep12(2019)097.

[500] O. J. C. Dias, H. S. Reall, and J. E. Santos. "Strong cosmic censorship for charged de Sitter black holes with a charged scalar field". *Class. Quantum Grav.* 36.4 (2019), p. 045005. ISSN: 1361-6382. DOI: 10.1088/1361-6382/aafcf2.

[501] R. Emparan and M. Tomašević. "Strong cosmic censorship in the BTZ black hole". *J. High Energy Phys.* 2020.6 (2020). ISSN: 1029-8479. DOI: 10.1007/jhep06(2020)038.

[502] T. McMaken. "Semiclassical instability of inner-extremal regular black holes". *Phys. Rev. D* 107 (12 2023), p. 125023. DOI: 10.1103/PhysRevD.107.125023.

[503] N. Zilberman, A. Levi, and A. Ori. "Quantum Fluxes at the Inner Horizon of a Spherical Charged Black Hole". *Phys. Rev. Lett.* 124 (17 2020), p. 171302. DOI: 10.1103/PhysRevLett.124.171302.

[504] H. Reeh and S. Schlieder. "Bemerkungen zur unitäräquivalenz von lorentzinvarianten feldern". *Il Nuovo Cimento (1955-1965)* 22 (1961), pp. 1051–1068. DOI: 10.1007/BF01645904.

[505] J. M. Renes. "Better bounds on optimal measurement and entanglement recovery, with applications to uncertainty and monogamy relations". *Phys. Rev. A* 96.4 (2017). ISSN: 2469-9934. DOI: 10.1103/physreva.96.042328.

[506] W. G. Unruh and R. M. Wald. "Information loss". *Reports on Progress in Physics* 80.9 (2017), p. 092002. ISSN: 1361-6633. DOI: 10.1088/1361-6633/aa778e.

[507] D. Marolf. "Unitarity and holography in gravitational physics". *Phys. Rev. D* 79 (2009), p. 044010. ISSN: 1550-2368. DOI: 10.1103/physrevd.79.044010.

[508] S. Raju. "Lessons from the information paradox". *Phys. Rep.* 943 (2022), pp. 1–80. DOI: 10.1016/j.physrep.2021.10.001.

[509] A. Kempf. "Replacing the Notion of Spacetime Distance by the Notion of Correlation". *Front. Phys.* 9 (2021), p. 655857. ISSN: 2296-424X. DOI: 10.3389/fphy.2021.655857.

[510] M. Saravani, S. Aslanbeigi, and A. Kempf. "Spacetime curvature in terms of scalar field propagators". *Phys. Rev. D* 93 (4 2016), p. 045026. DOI: 10.1103/PhysRevD.93.045026.

[511] A. Akil, M. Cadoni, L. Modesto, M. Oi, and A. P. Sanna. "Semiclassical spacetimes at super-Planckian scales from delocalized sources". *Phys. Rev. D* 108 (4 2023), p. 044051. DOI: 10.1103/PhysRevD.108.044051.

[512] T. Jacobson. "Boundary unitarity and the black hole information paradox". *Int. J. Modern Phys. D* 22.12 (2013), p. 1342002. ISSN: 1793-6594. DOI: 10.1142/s0218271813420029.

[513] T. Jacobson and P. Nguyen. "Diffeomorphism invariance and the black hole information paradox". *Phys. Rev. D* 100 (2019), p. 046002. ISSN: 2470-0029. DOI: 10.1103/physrevd.100.046002.

[514] K. Papadodimas and S. Raju. "Local Operators in the Eternal Black Hole". *Phys. Rev. Lett.* 115.21 (2015), p. 211601. DOI: 10.1103/PhysRevLett.115.211601. arXiv: 1502.06692 [hep-th].

[515] S. Raju. "A Toy Model of the Information Paradox in Empty Space". *SciPost Phys.* 6 (2019), p. 073. DOI: 10.21468/SciPostPhys.6.6.073.

[516] T. Chakraborty, J. Chakravarty, and P. Paul. "Monogamy paradox in empty flat space". *Phys. Rev.D* 106.8 (2022). ISSN: 2470-0029. DOI: 10.1103/physrevd.106.086002.

[517] C. Rovelli. "What Is Observable in Classical and Quantum Gravity?" *Class. Quant. Grav.* 8 (1991), pp. 297–316. DOI: 10.1088/0264-9381/8/2/011.

[518] J. D. Brown and K. V. Kuchař. "Dust as a standard of space and time in canonical quantum gravity". *Phys. Rev. D* 51 (1995), pp. 5600–5629. ISSN: 0556-2821. DOI: 10.1103/physrevd.51.5600.

[519] S. B. Giddings. "Gravitational dressing, soft charges, and perturbative gravitational splitting". *Phys. Rev. D* 100.12 (2019), p. 126001. DOI: 10.1103/PhysRevD.100.126001. arXiv: 1903.06160 [hep-th].

[520] S. B. Giddings and J. Perkins. "Perturbative quantum evolution of the gravitational state and dressing in general backgrounds". *Phys. Rev. D* 110.2 (2024), p. 026012. DOI: 10.1103/PhysRevD.110.026012. arXiv: 2209.06836 [hep-th].

[521] S. B. Giddings. "Quantum gravity observables: observation, algebras, and mathematical structure*". *J. Phys. A* 58.41 (2025), p. 415401. DOI: 10.1088/1751-8121/ae0b12. arXiv: 2505.22708 [hep-th].

[522] S. B. Giddings. *Gravitational dressing: from the crossed product to more general algebraic and mathematical structure*. 2025. arXiv: 2510.24833 [hep-th].

[523] W. Donnelly and S. B. Giddings. "Observables, gravitational dressing, and obstructions to locality and subsystems". *Phys. Rev. D* 94.10 (2016), p. 104038. DOI: 10.1103/PhysRevD.94.104038. arXiv: 1607.01025 [hep-th].

[524] W. Donnelly and S. B. Giddings. "Gravitational splitting at first order: Quantum information localization in gravity". *Phys. Rev. D* 98.8 (2018), p. 086006. DOI: 10.1103/PhysRevD.98.086006. arXiv: 1805.11095 [hep-th].

[525] S. B. Giddings, D. Marolf, and J. B. Hartle. "Observables in effective gravity". *Phys. Rev. D* 74 (2006), p. 064018. ISSN: 1550-2368. DOI: 10.1103/physrevd.74.064018.

[526] C. Goeller, P. A. Höhn, and J. Kirklin. *Diffeomorphism-invariant observables and dynamical frames in gravity: reconciling bulk locality with general covariance*. 2022. arXiv: 2206.01193 [hep-th].

[527] S. B. Giddings. "Quantum-First Gravity". *Found. Phys.* 49.3 (2019), pp. 177–190. ISSN: 1572-9516. DOI: 10.1007/s10701-019-00239-1.

[528] S. B. Giddings. *The unitarity crisis, nonviolent unitarization, and implications for quantum spacetime*. 2024. arXiv: 2412.18650 [hep-th].

[529] J. de Boer et al. *Frontiers of Quantum Gravity: shared challenges, converging directions*. 2022. arXiv: 2207.10618 [hep-th].

[530] G. 't Hooft. *Dimensional Reduction in Quantum Gravity*. 1993. arXiv: gr-qc/9310026 [gr-qc].

[531] L. Susskind. "The world as a hologram". *J. Math. Phys.* 36.11 (1995), pp. 6377–6396. ISSN: 1089-7658. DOI: 10.1063/1.531249.

[532] E. Witten. *Anti De Sitter Space And Holography*. 1998. arXiv: hep-th/9802150 [hep-th].

[533] J. Polchinski. *S-Matrices from AdS Spacetime*. 1999. arXiv: hep-th/9901076 [hep-th].

[534] L. Susskind. "Holography in the flat space limit". *AIP Conference Proceedings* 493.1 (1999), pp. 98–112. DOI: 10.1063/1.1301570.

[535] S. B. Giddings. "Flat-space scattering and bulk locality in the AdS-CFT correspondence". *Phys. Rev. D* 61 (10 2000), p. 106008. DOI: 10.1103/PhysRevD.61.106008.

[536] O. Aharony, S. S. Gubser, J. Maldacena, H. Ooguri, and Y. Oz. "Large N field theories, string theory and gravity". *Phys. Rep.* 323.3–4 (2000), pp. 183–386. ISSN: 0370-1573. DOI: 10.1016/s0370-1573(99)00083-6.

[537] J. de Boer and S. N. Solodukhin. "A holographic reduction of Minkowski space–time". *Nucl. Phys. B* 665 (2003), pp. 545–593. ISSN: 0550-3213. DOI: 10.1016/s0550-3213(03)00494-2.

[538] C. Dappiaggi. "BMS field theory and holography in asymptotically flat space-times". *J. High Energy Phys.* 2004.11 (2004), pp. 011–011. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2004/11/011.

[539] G. Arcioni and C. Dappiaggi. "Holography in asymptotically flat spacetimes and the BMS group". *Class. Quantum Gravity* 21.23 (2004), pp. 5655–5674. ISSN: 1361-6382. DOI: 10.1088/0264-9381/21/23/022.

[540] G. Arcioni and C. Dappiaggi. "Exploring the holographic principle in asymptotically flat spacetimes via the BMS group". *Nucl. Phys. B* 674.3 (2003), pp. 553–592. ISSN: 0550-3213. DOI: 10.1016/j.nuclphysb.2003.09.051.

[541] C. Dappiaggi, V. Moretti, and N. Pinamonti. "Rigorous steps towards holography in asymptotically flat spacetimes". *Rev. Math. Phys.* 18.04 (2006), pp. 349–415. ISSN: 1793-6659. DOI: 10.1142/s0129055x0600270x.

[542] D. Marolf. "Holography without strings?" *Class. Quantum Gravity* 31.1 (2013), p. 015008. ISSN: 1361-6382. DOI: 10.1088/0264-9381/31/1/015008.

[543] S. B. Giddings and A. Kinsella. "Gauge-invariant observables, gravitational dressings, and holography in AdS". *J. High Energy Phys.* 2018.11 (2018). ISSN: 1029-8479. DOI: 10.1007/jhep11(2018)074.

[544] S. B. Giddings. "Holography and unitarity". *J. High Energy Phys.* 2020.11 (2020). ISSN: 1029-8479. DOI: 10.1007/jhep11(2020)056.

[545] L. Donnay, A. Fiorucci, Y. Herfray, and R. Ruzziconi. "Carrollian Perspective on Celestial Holography". *Phys. Rev. Lett.* 129.7 (2022), p. 071602. ISSN: 1079-7114. DOI: 10.1103/physrevlett.129.071602.

[546] L. Donnay. *Celestial holography: An asymptotic symmetry perspective*. 2023. arXiv: 2310.12922 [hep-th].

[547] S. Pasterski. *A Chapter on Celestial Holography*. 2023. arXiv: 2310.04932 [hep-th].

[548] C. Chowdhury, V. Godet, O. Papadoulaki, and S. Raju. "Holography from the Wheeler–DeWitt equation". *J. High Energy Phys.* 2022.3 (2022). ISSN: 1029-8479. DOI: 10.1007/jhep03(2022)019.

[549] G. Araujo-Regado, R. Khan, and A. C. Wall. "Cauchy slice holography: a new AdS/CFT dictionary". *J. High Energy Phys.* 2023.3 (2023). ISSN: 1029-8479. DOI: 10.1007/jhep03(2023)026.

[550] L. Hui and I. Yang. "Complementarity plus backreaction is enough". *Phys. Rev. D* 89 (2014), p. 084011. ISSN: 1550-2368. DOI: 10.1103/physrevd.89.084011.

[551] J. Preskill. "Do black holes destroy information". *Proceedings of the International Symposium on Black Holes, Membranes, Wormholes and Superstrings, S. Kalara and DV Nanopoulos, eds.(World Scientific, Singapore, 1993) pp*. World Scientific. 1992, pp. 22–39. DOI: 10.48550/arXiv.hep-th/9209058.

[552] S. B. Giddings. "Black hole information, unitarity, and nonlocality". *Phys. Rev. D* 74.10 (2006), p. 106005. DOI: 10.1103/PhysRevD.74.106005.

[553] A. Puhm, F. Rojas, and T. Ugajin. "(Non-adiabatic) string creation on nice slices in Schwarzschild black holes". *J. High Energy Phys.* 2017.4 (2017). ISSN: 1029-8479. DOI: 10.1007/jhep04(2017)156.

[554] J. Polchinski. *String Theory and Black Hole Complementarity*. 1995. arXiv: hep-th/9507094 [hep-th].

[555] L. Hardy. "Towards quantum gravity: a framework for probabilistic theories with non-fixed causal structure". *J. Phys. A* 40 (2007), pp. 3081–3099. ISSN: 1751-8121. DOI: 10.1088/1751-8113/40/12/s12.

[556] G. Chiribella, G. M. D'Ariano, P. Perinotti, and B. Valiron. "Quantum computations without definite causal structure". *Phys. Rev. A* 88 (2 2013), p. 022318. DOI: 10.1103/PhysRevA.88.022318.

[557] E. Castro-Ruiz, F. Giacomini, and Č. Brukner. "Dynamics of Quantum Causal Structures". *Phys. Rev. X* 8 (1 2018), p. 011047. DOI: 10.1103/PhysRevX.8.011047.

[558] L. Parker and F. Costa. "Background Independence and Quantum Causal Structure". *Quantum* 6 (2022), p. 865. ISSN: 2521-327X. DOI: 10.22331/q-2022-11-28-865.

[559] A. Ashtekar and M. Bojowald. "Black hole evaporation: a paradigm". *Class. Quantum Gravity* 22.16 (2005), pp. 3349–3362. ISSN: 1361-6382. DOI: 10.1088/0264-9381/22/16/014.

[560] J. C. Schindler, A. Aguirre, and A. Kuttner. "Understanding black hole evaporation using explicitly computed Penrose diagrams". *Phys. Rev. D* 101 (2020), p. 024010. ISSN: 2470-0029. DOI: 10.1103/physrevd.101.024010.

[561] A. Ashtekar. "Black Hole Evaporation: A Perspective from Loop Quantum Gravity". *Universe* 6.2 (2020), p. 21. ISSN: 2218-1997. DOI: 10.3390/universe6020021.

[562] A. Ashtekar. "Black hole evaporation in loop quantum gravity". *Gen. Relativ. Gravit.* 57 (2025), p. 48. ISSN: 1572-9532. DOI: 10.1007/s10714-025-03380-7.

[563] W. A. Hiscock. "Models of evaporating black holes. I". *Phys. Rev. D* 23.12 (1981), p. 2813. DOI: 10.1103/PhysRevD.23.2813.

[564] W. A. Hiscock. "Models of evaporating black holes. II. Effects of the outgoing created radiation". *Phys. Rev. D* 23 (12 1981), pp. 2823–2827. DOI: 10.1103/PhysRevD.23.2823.

[565] S. A. Hayward. "Formation and Evaporation of Nonsingular Black Holes". *Phys. Rev. Lett.* 96.3 (2006). ISSN: 1079-7114. DOI: 10.1103/physrevlett.96.031103.

[566] R. Carballo-Rubio et al. "Towards a non-singular paradigm of black hole physics". *J. Cosmology Astropart. Phys.* 05 (2025), p. 003. DOI: 10.1088/1475-7516/2025/05/003. arXiv: 2501.05505 [gr-qc].

[567] L. C. Barbado, C. Barceló, and L. J. Garay. "Hawking radiation as perceived by different observers". *Class. Quantum Grav.* 28.12 (2011), p. 125021. ISSN: 1361-6382. DOI: 10.1088/0264-9381/28/12/125021.

[568] L. C. Barbado, C. Barceló, L. J. Garay, and G. Jannes. "Hawking versus Unruh effects, or the difficulty of slowly crossing a black hole horizon". *J. High Energy Phys.* 2016.10 (2016). ISSN: 1029-8479. DOI: 10.1007/jhep10(2016)161.

[569] P. M. Alsing and I. Fuentes. "Observer-dependent entanglement". *Class. Quantum Grav.* 29.22 (2012), p. 224001. ISSN: 1361-6382. DOI: 10.1088/0264-9381/29/22/224001.

[570] J. D. Vuyst, S. Eccles, P. A. Höhn, and J. Kirklin. *Gravitational entropy is observer-dependent*. 2025. arXiv: 2405.00114 [hep-th].

[571] M. Smerlak and S. Singh. "New perspectives on Hawking radiation". *Phys. Rev. D* 88 (10 2013), p. 104023. DOI: 10.1103/PhysRevD.88.104023.

[572] K. K. Ng, C. Zhang, J. Louko, and R. B. Mann. "A little excitement across the horizon". *New. J. Phys.* 24.10 (2022), p. 103018. DOI: 10.1088/1367-2630/ac9547.

[573] M. R. Preciado-Rivas, M. Naeem, R. B. Mann, and J. Louko. "More excitement across the horizon". *Phys. Rev. D* 110 (2 2024), p. 025002. DOI: 10.1103/PhysRevD.110.025002.

[574] T. Dray and G. 't Hooft. "The Effect of Spherical Shells of Matter on the Schwarzschild Black Hole". *Commun. Math. Phys.* 99 (1985), pp. 613–625. DOI: 10.1007/BF01215912.

[575] G. 't Hooft. "Strings from gravity". *Physica Scripta* 1987.T15 (1987), p. 143. DOI: 10.1088/0031-8949/1987/T15/019.

[576] N. Gaddam and N. Groenenboom. "Soft graviton exchange and the information paradox". *Phys. Rev.D* 109.2 (2024), p. 026007. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.109.026007.

[577] V. Ferrari, P. Pendenza, and G. Veneziano. "Beamlike Gravitational Waves and Their Geodesics". *Gen. Rel. Grav.* 20 (1988), pp. 1185–1191. DOI: 10.1007/BF00758938.

[578] K. Sfetsos. "On gravitational shock waves in curved spacetimes". *Nucl. Phys. B* 436.3 (1995), pp. 721–745. ISSN: 0550-3213. DOI: 10.1016/0550-3213(94)00573-w.

[579] S. H. Shenker and D. Stanford. "Black holes and the butterfly effect". *J. High Energy Phys.* 2014.3 (2014), pp. 1–25.

[580] P. Betzios, N. Gaddam, and O. Papadoulaki. "The Black Hole S-matrix from Quantum Mechanics". *J. High Energy Phys.* 11 (2016), p. 131. DOI: 10.1007/JHEP11(2016)131. arXiv: 1607.07885 [hep-th].

[581] N. Gaddam and N. Groenenboom. "2 $\rightarrow$ 2N scattering: Eikonalisation and the Page curve". *J. High Energy Phys.* 01 (2022), p. 146. DOI: 10.1007/JHEP01(2022)146. arXiv: 2110.14673 [hep-th].

[582] Y. BenTov and J. Swearngin. "Gravitational shockwaves on rotating black holes". *Gen. Rel. Grav.* 51.2 (2019), p. 25. DOI: 10.1007/s10714-019-2512-7. arXiv: 1706.03430 [gr-qc].

[583] J. Polchinski. *Chaos in the black hole S-matrix*. 2015. arXiv: 1505.08108 [hep-th].

[584] G. 't Hooft. "Graviton dominance in ultra-high-energy scattering". *Phys. Lett. B* 198.1 (1987), pp. 61–63. ISSN: 0370-2693. DOI: 10.1016/0370-2693(87)90159-6.

[585] H. Verlinde and E. Verlinde. "Scattering at planckian energies". *Nucl. Phys. B* 371.1–2 (1992), pp. 246–268. ISSN: 0550-3213. DOI: 10.1016/0550-3213(92)90236-5.

[586] N. Gaddam, N. Groenenboom, and G. 't Hooft. "Quantum gravity on the black hole horizon". *J. High Energy Phys.* 01 (2022), p. 023. DOI: 10.1007/JHEP01(2022)023. arXiv: 2012.02357 [hep-th].

[587] P. D. Vecchia, C. Heissenberg, R. Russo, and G. Veneziano. *The gravitational eikonal: from particle, string and brane collisions to black-hole encounters*. 2024. arXiv: 2306.16488 [hep-th].

[588] D. Amati, M. Ciafaloni, and G. Veneziano. "Towards an S-matrix description of gravitational collapse". *J. High Energy Phys.* 2008.02 (2008), pp. 049–049. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/02/049.

[589] G. Veneziano and J. Wosiek. "Exploring an S-matrix for gravitational collapse". *J. High Energy Phys.* 2008.09 (2008), pp. 023–023. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/09/023.

[590] G. Veneziano and J. Wosiek. "Exploring an S-matrix for gravitational collapse II: a momentum space analysis". *J. High Energy Phys.* 2008.09 (2008), pp. 024–024. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/09/024.

[591] G. Marchesini and E. Onofri. "High energy gravitational scattering: a numerical study". *J. High Energy Phys.* 2008.06 (2008), pp. 104–104. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/06/104.

[592] D. Kabat and M. Ortiz. "Eikonal quantum gravity and planckian scattering". *Nucl. Phys. B* 388.2 (1992), pp. 570–592. ISSN: 0550-3213. DOI: 10.1016/0550-3213(92)90627-n.

[593] P. D. Vecchia, C. Heissenberg, R. Russo, and G. Veneziano. "Universality of ultra-relativistic gravitational scattering". *Phys. Lett. B* 811 (2020), p. 135924. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2020.135924.

[594] M. Ciafaloni and D. Colferai. "S-matrix and quantum tunneling in gravitational collapse". *J. High Energy Phys.* 2008.11 (2008), pp. 047–047. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/11/047.

[595] G. Dvali, C. Gomez, R. Isermann, D. Lüst, and S. Stieberger. "Black hole formation and classicalization in ultra-Planckian scattering". *Nucl. Phys. B* 893 (2015), pp. 187–235. ISSN: 0550-3213. DOI: 10.1016/j.nuclphysb.2015.02.004.

[596] A. Addazi, M. Bianchi, and G. Veneziano. "Glimpses of black hole formation/evaporation in highly inelastic, ultra-planckian string collisions". *J. High Energy Phys.* 2017.2 (2017). ISSN: 1029-8479. DOI: 10.1007/jhep02(2017)111.

[597] M. Ciafaloni, D. Colferai, and G. Veneziano. "Infrared features of gravitational scattering and radiation in the eikonal approach". *Phys. Rev. D* 99 (6 2019), p. 066008. DOI: 10.1103/PhysRevD.99.066008.

[598] T. Banks and W. Fischler. *A Model for High Energy Scattering in Quantum Gravity*. 1999. arXiv: hep-th/9906038 [hep-th].

[599] O. Bertolami, J. Paramos, and S. G. Turyshev. *General Theory of Relativity: Will it survive the next decade?* 2006. arXiv: gr-qc/0602016 [gr-qc].

[600] L. Savrov. "Gravitational shielding and the equivalence principle". *Gravitation and Cosmology* 18.4 (2012), pp. 270–278. DOI: 10.1134/S0202289312040081.

[601] A. Dukehart and D. Mattingly. "Diffeomorphism invariance and quantum mechanical paradoxes". *Int. J. Mod. Phys. D* 33.01 (2024), p. 2350109. DOI: 10.1142/S0218271823501092.

[602] D. L. Danielson, G. Satishchandran, and R. M. Wald. "Black holes decohere quantum superpositions". *Int. J. Mod. Phys. D* 31.14 (2022), p. 2241003. DOI: 10.1142/S0218271822410036. arXiv: 2205.06279 [hep-th].

[603] D. L. Danielson, G. Satishchandran, and R. M. Wald. "Local description of decoherence of quantum superpositions by black holes and other bodies". *Phys. Rev. D* 111.2 (2025), p. 025014. DOI: 10.1103/PhysRevD.111.025014. arXiv: 2407.02567 [hep-th].

[604] A. Belenchia, R. M. Wald, F. Giacomini, E. Castro-Ruiz, Č. Brukner, and M. Aspelmeyer. "Quantum superposition of massive objects and the quantization of gravity". *Phys. Rev. D* 98 (2018), p. 126009. ISSN: 2470-0029. DOI: 10.1103/physrevd.98.126009.

[605] D. L. Danielson and G. Satishchandran. *Horizons and Soft Quantum Information*. 2025. arXiv: 2512.20754 [hep-th].

[606] P. Hájíček, B. S. Kay, and K. V. Kuchař. "Quantum collapse of a self-gravitating shell: Equivalence to Coulomb scattering". *Phys. Rev. D* 46 (12 1992), pp. 5439–5448. DOI: 10.1103/PhysRevD.46.5439.

[607] P. Hájíček. "Quantum mechanics of gravitational collapse". *Commun. Math. Phys.* 150.3 (1992), pp. 545–559. DOI: 10.1007/BF02096961.

[608] P. Hájíček. "Unitary dynamics of spherical null gravitating shells". *Nucl. Phys. B* 603.3 (2001), pp. 555–577. ISSN: 0550-3213. DOI: 10.1016/S0550-3213(01)00140-7.

[609] J. Louko, B. F. Whiting, and J. L. Friedman. "Hamiltonian spacetime dynamics with a spherical null-dust shell". *Phys. Rev.D* 57.4 (1998), pp. 2279–2298. ISSN: 0556-2821, 1089-4918. DOI: 10.1103/PhysRevD.57.2279.

[610] A. Corichi, G. Cruz-Pacheco, A. Minzoni, P. Padilla, M. Rosenbaum, M. P. Ryan, N. F. Smyth, and T. Vukasinac. "Quantum collapse of a small dust shell". *Phys. Rev. D* 65 (6 2002), p. 064006. DOI: 10.1103/PhysRevD.65.064006.

[611] C. Vaz, S. Gutti, C. Kiefer, and T. P. Singh. "Quantum gravitational collapse and Hawking radiation in $2+1$ dimensions". *Phys. Rev. D* 76 (12 2007), p. 124021. DOI: 10.1103/PhysRevD.76.124021.

[612] C. Vaz. "Quantum collapse of a thin shell revisited". *Phys. Rev. D* 105 (2022), p. 086020. ISSN: 2470-0029. DOI: 10.1103/physrevd.105.086020.

[613] V. Baccetti, S. Murk, and D. R. Terno. "Black hole evaporation and semiclassical thin shell collapse". *Phys. Rev. D* 100 (6 2019), p. 064054. DOI: 10.1103/PhysRevD.100.064054.

[614] E. Gaztañaga, K. S. Kumar, S. Pradhan, and M. Gabler. "Gravitational bounce from the quantum exclusion principle". *Phys. Rev. D* 111 (10 2025), p. 103537. DOI: 10.1103/PhysRevD.111.103537.

[615] V. Berezin. "Quantum black hole model and Hawking's radiation". *Phys. Rev. D* 55 (4 1997), pp. 2139–2151. DOI: 10.1103/PhysRevD.55.2139.

[616] D. Malafarina. "Classical Collapse to Black Holes and Quantum Bounces: A Review". *Universe* 3.2 (2017), p. 48. ISSN: 2218-1997. DOI: 10.3390/universe3020048.

[617] J. Ziprick, J. Gegenberg, and G. Kunstatter. "Polymer quantization of a self-gravitating thin shell". *Phys. Rev. D* 94 (10 2016), p. 104076. DOI: 10.1103/PhysRevD.94.104076.

[618] R. Penrose. *unpublished*. 1974.

[619] H. Matschull. "Black hole creation in 2 + 1 dimensions". *Class. Quantum Gravity* 16.3 (1999), pp. 1069–1095. ISSN: 1361-6382. DOI: 10.1088/0264-9381/16/3/032.

[620] D. M. Eardley and S. B. Giddings. "Classical black hole production in high-energy collisions". *Phys. Rev. D* 66 (2002), p. 044011. ISSN: 1089-4918. DOI: 10.1103/physrevd.66.044011.

[621] V. S. Rychkov. "Black hole production in particle collisions and higher curvature gravity". *Phys. Rev. D* 70 (4 2004), p. 044003. DOI: 10.1103/PhysRevD.70.044003.

[622] H. Yoshino and V. S. Rychkov. "Improved analysis of black hole formation in high-energy particle collisions". *Phys. Rev. D* 71 (10 2005), p. 104028. DOI: 10.1103/PhysRevD.71.104028.

[623] S. B. Giddings and V. S. Rychkov. "Black holes from colliding wavepackets". *Phys. Rev. D* 70 (10 2004), p. 104026. DOI: 10.1103/PhysRevD.70.104026.

[624] S. D. Hsu. "Quantum production of black holes". *Phys. Lett. B* 555.1–2 (2003), pp. 92–98. ISSN: 0370-2693. DOI: 10.1016/s0370-2693(03)00012-1.

[625] V. Cardoso, L. Gualtieri, C. Herdeiro, and U. Sperhake. "Exploring New Physics Frontiers Through Numerical Relativity". *Living Rev. Relativ.* 18 (2015), p. 1. ISSN: 1433-8351. DOI: 10.1007/lrr-2015-1.

[626] M. W. Choptuik and F. Pretorius. "Ultrarelativistic Particle Collisions". *Phys. Rev. Lett.* 104 (2010), p. 111101. ISSN: 1079-7114. DOI: 10.1103/physrevlett.104.111101.

[627] U. Sperhake, V. Cardoso, F. Pretorius, E. Berti, T. Hinderer, and N. Yunes. "Cross Section, Final Spin, and Zoom-Whirl Behavior in High-Energy Black-Hole Collisions". *Phys. Rev. Lett.* 103 (13 2009), p. 131102. DOI: 10.1103/PhysRevLett.103.131102.

[628] P. D. D'Eath and P. N. Payne. "Gravitational radiation in black-hole collisions at the speed of light. I. Perturbation treatment of the axisymmetric collision". *Phys. Rev. D* 46 (2 1992), pp. 658–674. DOI: 10.1103/PhysRevD.46.658.

[629] M. Ciafaloni, D. Colferai, and G. Veneziano. "Emerging Hawking-Like Radiation from Gravitational Bremsstrahlung Beyond the Planck Scale". *Phys. Rev. Lett.* 115 (17 2015), p. 171301. DOI: 10.1103/PhysRevLett.115.171301.

[630] A. Addazi, M. Bianchi, and G. Veneziano. "Soft gravitational radiation from ultra-relativistic collisions at sub- and sub-sub-leading order". *J. High Energy Phys.* 2019 (2019), p. 50. ISSN: 1029-8479. DOI: 10.1007/jhep05(2019)050.

[631] Z. Bern, J. Parra-Martinez, R. Roiban, M. S. Ruf, C. Shen, M. P. Solon, and M. Zeng. "Scattering Amplitudes, the Tail Effect, and Conservative Binary Dynamics at $O(G^4)$". *Phys. Rev. Lett.* 128 (2022), p. 161103. ISSN: 1079-7114. DOI: 10.1103/physrevlett.128.161103.

[632] T. Damour. "High-energy gravitational scattering and the general relativistic two-body problem". *Phys. Rev. D* 97 (4 2018), p. 044038. DOI: 10.1103/PhysRevD.97.044038.

[633] D. Neill and I. Z. Rothstein. "Classical space–times from the S-matrix". *Nucl. Phys. B* 877.2 (2013), pp. 177–189. ISSN: 0550-3213. DOI: 10.1016/j.nuclphysb.2013.09.007.

[634] H. Yoshino and Y. Nambu. "Black hole formation in the grazing collision of high-energy particles". *Phys. Rev. D* 67 (2 2003), p. 024009. DOI: 10.1103/PhysRevD.67.024009.

[635] W. E. East and F. Pretorius. "Ultrarelativistic Black Hole Formation". *Phys. Rev. Lett.* 110 (10 2013), p. 101101. DOI: 10.1103/PhysRevLett.110.101101.

[636] T. Anous, T. Hartman, A. Rovai, and J. Sonner. "Black hole collapse in the 1/c expansion". *J. High Energy Phys.* 2016 (2016), p. 7. ISSN: 1029-8479. DOI: 10.1007/jhep07(2016)123.

[637] F. M. Haehl and Y. Zhao. "Operator growth and black hole formation". *J. High Energy Phys.* 2023 (2023), p. 7. ISSN: 1029-8479. DOI: 10.1007/jhep07(2023)184.

[638] I. Kourkoulou and J. Maldacena. *Pure states in the SYK model and nearly-AdS_2 gravity*. 2017. arXiv: 1707.02325 [hep-th].

[639] C. Barceló, S. Liberati, and M. Visser. "Analogue Gravity". *Liv. Rev. Relativ.* 8.1 (2005), p. 12. ISSN: 1433-8351. DOI: 10.12942/lrr-2005-12.

[640] W. G. Unruh. "Sonic analogue of black holes and the effects of high frequencies on black hole evaporation". *Phys. Rev. D* 51 (6 1995), pp. 2827–2838. DOI: 10.1103/PhysRevD.51.2827.

[641] J. Steinhauer. "Observation of quantum Hawking radiation and its entanglement in an analogue black hole". *Nat. Phys.* 12.10 (2016), pp. 959–965. DOI: 10.1038/nphys3863.

[642] O. Lahav, A. Itah, A. Blumkin, C. Gordon, S. Rinott, A. Zayats, and J. Steinhauer. "Realization of a Sonic Black Hole Analog in a Bose-Einstein Condensate". *Phys. Rev. Lett.* 105 (24 2010), p. 240401. DOI: 10.1103/PhysRevLett.105.240401.

[643] J. Drori, Y. Rosenberg, D. Bermudez, Y. Silberberg, and U. Leonhardt. "Observation of Stimulated Hawking Radiation in an Optical Analogue". *Phys. Rev. Lett.* 122 (1 2019), p. 010404. DOI: 10.1103/PhysRevLett.122.010404.

[644] J. Steinhauer. "Observation of self-amplifying Hawking radiation in an analogue black-hole laser". *Nat. Phys.* 10.11 (2014), pp. 864–869. ISSN: 1745-2481. DOI: 10.1038/nphys3104.

[645] S. Bose, I. Fuentes, A. A. Geraci, S. M. Khan, S. Qvarfort, M. Rademacher, M. Rashid, M. Toroš, H. Ulbricht, and C. C. Wanjura. "Massive quantum systems as interfaces of quantum mechanics and gravity". *Rev. Mod. Phys.* 97 (1 2025), p. 015003. DOI: 10.1103/RevModPhys.97.015003.

[646] J. R. Muñoz de Nova, K. Golubkov, V. I. Kolobov, and J. Steinhauer. "Observation of thermal Hawking radiation and its temperature in an analogue black hole". *Nature* 569.7758 (2019), pp. 688–691. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1241-0.

[647] S. Liberati, G. Tricella, and A. Trombettoni. "The information loss problem: an analogue gravity perspective". *Entropy* 21.10 (2019), p. 940. DOI: 10.3390/e21100940.

[648] D. Chowdhury, A. Georges, O. Parcollet, and S. Sachdev. "Sachdev-Ye-Kitaev models and beyond: Window into non-Fermi liquids". *Rev. Modern Phys.* 94.3 (2022), p. 035004. DOI: 10.1103/revmodphys.94.035004.

[649] L. J. Henderson, R. A. Hennigar, R. B. Mann, A. R. Smith, and J. Zhang. "Harvesting entanglement from the black hole vacuum". *Class. Quantum Grav.* 35.21 (2018), 21LT02. DOI: 10.1088/1361-6382/aae27e.

[650] K. K. Ng, L. Hodgkinson, J. Louko, R. B. Mann, and E. Martín-Martínez. "Unruh-DeWitt detector response along static and circular-geodesic trajectories for Schwarzschild–anti-de Sitter black holes". *Phys. Rev. D* 90 (6 2014), p. 064003. DOI: 10.1103/PhysRevD.90.064003.

[651] D. Bhattacharya, J. Louko, and R. B. Mann. *Probing hidden topology with quantum detectors*. 2025. DOI: 10.1103/PhysRevD.111.045005.

[652] A. R. H. Smith and R. B. Mann. "Looking inside a black hole". *Class. Quantum Grav.* 31.8 (2014), p. 082001. ISSN: 1361-6382. DOI: 10.1088/0264-9381/31/8/082001.

[653] L. Hodgkinson and J. Louko. "Static, stationary, and inertial Unruh-DeWitt detectors on the BTZ black hole". *Phys. Rev. D* 86 (6 2012), p. 064031. DOI: 10.1103/PhysRevD.86.064031.

[654] L. Hodgkinson. *Particle detectors in curved spacetime quantum field theory*. 2013. arXiv: 1309.7281 [gr-qc].

[655] J. Paczos and L. C. Barbado. "Hawking radiation for detectors in superposition of locations outside a black hole". *Phys. Rev. D* 108.12 (2023), p. 125015. DOI: 10.1103/PhysRevD.108.125015.

[656] L. J. Henderson, R. A. Hennigar, R. B. Mann, A. R. Smith, and J. Zhang. "Anti-Hawking phenomena". *Phys. Lett.B* 809 (2020), p. 135732. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2020.135732.

[657] L. J. Henderson, S. Y. Ding, and R. B. Mann. "Entanglement harvesting with a twist". *AVS Quantum Science* 4.1 (2022), p. 014402. ISSN: 2639-0213. DOI: 10.1116/5.0078314.

[658] B. A. Juárez-Aubry and J. Louko. "Onset and decay of the 1 + 1 Hawking-Unruh effect: what the derivative-coupling detector saw". *Class. Quantum Grav.* 31.24 (2014), p. 245007. ISSN: 1361-6382. DOI: 10.1088/0264-9381/31/24/245007.

[659] S. Wang, M. R. Preciado-Rivas, M. Spadafora, and R. B. Mann. "Singular excitement beyond the horizon of a rotating black hole". *Phys. Rev. D* 110.6 (2024), p. 065013. DOI: 10.1103/PhysRevD.110.065013.

[660] M. Spadafora, M. Naeem, M. R. Preciado-Rivas, R. B. Mann, and J. Louko. *Deep in the Knotted Black Hole*. 2024. arXiv: 2412.02755 [gr-qc].

[661] J. Foo, S. Onoe, and M. Zych. "Unruh-DeWitt detectors in quantum superpositions of trajectories". *Phys. Rev. D* 102 (8 2020), p. 085013. DOI: 10.1103/PhysRevD.102.085013.

[662] J. Foo, S. Onoe, R. B. Mann, and M. Zych. "Thermality, causality, and the quantum-controlled Unruh–deWitt detector". *Phys. Rev. Res.* 3 (4 2021), p. 043056. DOI: 10.1103/PhysRevResearch.3.043056.

[663] J. Foo, R. B. Mann, and M. Zych. "Entanglement amplification between superposed detectors in flat and curved spacetimes". *Phys. Rev. D* 103 (6 2021), p. 065013. DOI: 10.1103/PhysRevD.103.065013.

[664] J. Foo, C. S. Arabaci, M. Zych, and R. B. Mann. "Quantum superpositions of Minkowski spacetime". *Phys. Rev. D* 107.4 (2023), p. 045014. DOI: 10.1103/PhysRevD.107.045014.

[665] L. Goel, E. A. Patterson, M. R. Preciado-Rivas, M. Torabian, R. B. Mann, and N. Afshordi. *Accelerated detector in a superposed spacetime*. 2024. arXiv: 2409.06818 [gr-qc].

[666] C. Suryaatmadja, C. S. Arabaci, M. P. G. Robbins, J. Foo, M. Zych, and R. B. Mann. *Signatures of Rotating Black Holes in Quantum Superposition*. 2023. arXiv: 2310.10864 [gr-qc].

[667] L. Niermann and L. C. Barbado. *Particle detectors in superposition in de Sitter spacetime*. 2024. arXiv: 2403.02087 [gr-qc].

[668] N. Stritzelberger, L. J. Henderson, V. Baccetti, N. C. Menicucci, and A. Kempf. "Entanglement harvesting with coherently delocalized matter". *Phys. Rev. D* 103 (1 2021), p. 016007. DOI: 10.1103/PhysRevD.103.016007.

[669] M. Hotta, A. Kempf, E. Martín-Martínez, T. Tomitsuka, and K. Yamaguchi. "Duality in the dynamics of Unruh-DeWitt detectors in conformally related spacetimes". *Phys. Rev. D* 101 (8 2020), p. 085017. DOI: 10.1103/PhysRevD.101.085017.

[670] N. Stritzelberger and A. Kempf. "Coherent delocalization in the light-matter interaction". *Phys. Rev. D* 101 (3 2020), p. 036007. DOI: 10.1103/PhysRevD.101.036007.

[671] W. Cong, C. Qian, M. R. Good, and R. B. Mann. "Effects of horizons on entanglement harvesting". *J. High Energy Phys.* 2020 (2020), p. 10. ISSN: 1029-8479. DOI: 10.1007/jhep10(2020)067.

[672] K. K. Ng, R. B. Mann, and E. Martín-Martínez. "New techniques for entanglement harvesting in flat and curved spacetimes". *Phys. Rev. D* 97 (12 2018), p. 125011. DOI: 10.1103/PhysRevD.97.125011.

[673] K. Gallock-Yoshimura, E. Tjoa, and R. B. Mann. "Harvesting entanglement with detectors freely falling into a black hole". *Phys. Rev. D* 104 (2 2021), p. 025001. DOI: 10.1103/PhysRevD.104.025001.

[674] D. Mendez-Avalos, L. J. Henderson, K. Gallock-Yoshimura, and R. B. Mann. "Entanglement harvesting of three Unruh-DeWitt detectors". *General Relativity and Gravitation* 54.8 (2022), p. 87. ISSN: 1572-9532. DOI: 10.1007/s10714-022-02956-x.

[675] E. Tjoa and R. B. Mann. "Harvesting correlations in Schwarzschild and collapsing shell spacetimes". *J. High Energy Phys.* 2020.8 (2020), p. 155. ISSN: 1029-8479. DOI: 10.1007/jhep08(2020)155.

[676] M. P. G. Robbins, L. J. Henderson, and R. B. Mann. *Entanglement Amplification from Rotating Black Holes*. 2020. arXiv: 2010.14517 [hep-th].

[677] A. Chakraborty, L. Hackl, and M. Zych. *Entanglement harvesting in quantum superposed spacetime*. 2024. arXiv: 2412.15870 [gr-qc].

[678] B. Šoda, V. Sudhir, and A. Kempf. "Acceleration-induced effects in stimulated light-matter interactions". *Phys. Rev. Lett.* 128.16 (2022), p. 163603. DOI: 10.1103/PhysRevLett.128.163603.

[679] Y. Pan and B. Zhang. "Gravity-induced transparency". *Phys. Rev. D* 109.12 (2024), p. 125018. DOI: 10.1103/PhysRevD.109.125018.

[680] L.-Q. Chen and F. Giacomini. *Quantum effects in gravity beyond the Newton potential from a delocalised quantum source*. 2024. arXiv: 2402.10288 [quant-ph].

[681] L.-Q. Chen, F. Giacomini, and C. Rovelli. "Quantum states of fields for quantum split sources". *Quantum* 7 (2023), p. 958. DOI: 10.22331/q-2023-03-20-958.

[682] C. Rovelli. "Quantum mechanics without time: A model". *Phys. Rev.D* 42.8 (1990), pp. 2638–2646. ISSN: 0556-2821. DOI: 10.1103/PhysRevD.42.2638.

[683] C. J. Fewster, D. W. Janssen, L. D. Loveridge, K. Rejzner, and J. Waldron. "Quantum Reference Frames, Measurement Schemes and the Type of Local Algebras in Quantum Field Theory". *Commun. Math. Phys.* 406 (2024), p. 19. ISSN: 1432-0916. DOI: 10.1007/s00220-024-05180-7.

[684] S. A. Ahmad, W. Chemissany, M. S. Klinger, and R. G. Leigh. "Quantum reference frames from top-down crossed products". *Phys. Rev. D* 110 (6 2024), p. 065003. DOI: 10.1103/PhysRevD.110.065003.

[685] S. Carrozza, S. Eccles, and P. A. Höhn. "Edge modes as dynamical frames: charges from post-selection in generally covariant theories". *Scipost Phys.* 17 (2024), p. 048. ISSN: 2542-4653. DOI: 10.21468/scipostphys.17.2.048.

[686] A. R. Brown, H. Gharibyan, G. Penington, and L. Susskind. *The Python's Lunch: geometric obstructions to decoding Hawking radiation*. 2019. arXiv: 1912.00228 [hep-th].

[687] A. R. Brown, L. Susskind, and Y. Zhao. "Quantum complexity and negative curvature". *Phys. Rev. D* 95 (2017), p. 045010. ISSN: 2470-0029. DOI: 10.1103/physrevd.95.045010.

[688] S. Baiguera, V. Balasubramanian, P. Caputa, S. Chapman, J. Haferkamp, M. P. Heller, and N. Y. Halpern. "Quantum complexity in gravity, quantum field theory, and quantum information science". *Phys. Rep.* 1159 (2026), pp. 1–77. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2025.11.001.

[689] N. Engelhardt and H. Liu. *Algebraic ER=EPR and Complexity Transfer*. 2023. arXiv: 2311.04281 [hep-th].

[690] A. Belin, R. C. Myers, S.-M. Ruan, G. Sárosi, and A. J. Speranza. "Does Complexity Equal Anything?" *Phys. Rev. Lett.* 128 (2022), p. 081602. ISSN: 1079-7114. DOI: 10.1103/physrevlett.128.081602.

[691] A. Almheiri. *Holographic Quantum Error Correction and the Projected Black Hole Interior*. 2018. arXiv: 1810.02055 [hep-th].

[692] E. Verlinde and H. Verlinde. "Black hole entanglement and quantum error correction". *J. High Energy Phys.* 2013.10 (2013). ISSN: 1029-8479. DOI: 10.1007/jhep10(2013)107.

[693] P. Hayden and G. Penington. "Learning the Alpha-bits of black holes". *J. High Energy Phys.* 2019.12 (2019). ISSN: 1029-8479. DOI: 10.1007/jhep12(2019)007.

[694] A. Almheiri, X. Dong, and D. Harlow. "Bulk locality and quantum error correction in AdS/CFT". *J. High Energy Phys.* 2015 (2015), p. 163. ISSN: 1029-8479. DOI: 10.1007/jhep04(2015)163.

[695] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill. "Holographic quantum error-correcting codes: toy models for the bulk/boundary correspondence". *J. High Energy Phys.* 2015 (2015), p. 149. ISSN: 1029-8479. DOI: 10.1007/jhep06(2015)149.

[696] J. Cotler, N. Hunter-Jones, J. Liu, and B. Yoshida. "Chaos, complexity, and random matrices". *J. High Energy Phys.* 2017 (2017), p. 48. ISSN: 1029-8479. DOI: 10.1007/jhep11(2017)048.

[697] T. Rivlin, S. Engineer, and V. Baumann. *Emergence of Classicality in Wigner's Friend Scenarios*. 2025. arXiv: 2507.21221 [quant-ph].

[698] W. H. Zurek et al. "Decoherence and the transition from quantum to classical". *Phys. today* 44.10 (1991), pp. 36–44.

[699] W. H. Zurek. "Decoherence, einselection, and the quantum origins of the classical". *Rev. Modern Phys.* 75.3 (2003), p. 715.

[700] W. H. Zurek. "Quantum Darwinism". *Nat. Phys.* 5.3 (2009), pp. 181–188. DOI: 10.1038/nphys1202. arXiv: 0903.5082 [quant-ph].

[701] F. G. S. L. Brandão, M. Piani, and P. Horodecki. "Generic emergence of classical features in quantum Darwinism". *Nat. Commun.* 6.1 (2015). ISSN: 2041-1723. DOI: 10.1038/ncomms8908.

[702] A. E. Allahverdyan, R. Balian, and T. M. Nieuwenhuizen. "Understanding quantum measurement from the solution of dynamical models". *Phys. Rep.* 525.1 (2013), pp. 1–166. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2012.11.001.

[703] J. Kofler and Č. Brukner. "Classical World Arising out of Quantum Physics under the Restriction of Coarse-Grained Measurements". *Phys. Rev. Lett.* 99.18 (2007). ISSN: 1079-7114. DOI: 10.1103/physrevlett.99.180403.

[704] F. Bibak, C. Cepollaro, N. M. Sánchez, B. Dakić, and Č. Brukner. *The classical limit of quantum mechanics through coarse-grained measurements*. 2025. arXiv: 2503.15642 [quant-ph].

[705] R. Omnès. *The Interpretation of Quantum Mechanics*. Princeton University Press, 1994.

[706] M. Gell-Mann and J. B. Hartle. "Classical equations for quantum systems". *Phys. Rev. D* 47 (8 1993), pp. 3345–3382. DOI: 10.1103/PhysRevD.47.3345.

[707] F. Lombardo and F. D. Mazzitelli. "Coarse graining and decoherence in quantum field theory". *Phys. Rev. D* 53.4 (1996), pp. 2001–2011. ISSN: 1089-4918. DOI: 10.1103/physrevd.53.2001.

[708] J. J. Halliwell. "Decoherent histories and hydrodynamic equations". *Phys. Rev. D* 58.10 (1998). ISSN: 1089-4918. DOI: 10.1103/physrevd.58.105015.

[709] W. H. Zurek and J. P. Paz. "Decoherence, chaos, and the second law". *Phys. Rev. Lett.* 72 (16 1994), pp. 2508–2511. DOI: 10.1103/PhysRevLett.72.2508.

[710] P. Strasberg. "Classicality with(out) decoherence: Concepts, relation to Markovianity, and a random matrix theory approach". *SciPost Phys.* 15.1 (2023). ISSN: 2542-4653. DOI: 10.21468/scipostphys.15.1.024.

[711] S. Choi, Y. Bao, X.-L. Qi, and E. Altman. "Quantum Error Correction in Scrambling Dynamics and Measurement-Induced Phase Transition". *Phys. Rev. Lett.* 125.3 (2020). ISSN: 1079-7114. DOI: 10.1103/physrevlett.125.030505.

[712] M. J. Gullans and D. A. Huse. "Dynamical Purification Phase Transition Induced by Quantum Measurements". *Phys. Rev. X* 10.4 (2020). ISSN: 2160-3308. DOI: 10.1103/physrevx.10.041020.

[713] R. Suzuki, J. Haferkamp, J. Eisert, and P. Faist. "Quantum complexity phase transitions in monitored random circuits". *Quantum* 9 (2025), p. 1627. ISSN: 2521-327X. DOI: 10.22331/q-2025-02-10-1627.

[714] B. Skinner, J. Ruhman, and A. Nahum. "Measurement-Induced Phase Transitions in the Dynamics of Entanglement". *Phys. Rev. X* 9.3 (2019). ISSN: 2160-3308. DOI: 10.1103/physrevx.9.031009.

[715] Y. Li, X. Chen, and M. P. A. Fisher. "Measurement-driven entanglement transition in hybrid quantum circuits". *Phys. Rev. B* 100.13 (2019). ISSN: 2469-9969. DOI: 10.1103/physrevb.100.134306.

[716] J. D. Bekenstein. "The quantum mass spectrum of the Kerr black hole". *JACOB BEKENSTEIN: The Conservative Revolutionary*. World Scientific, 2020, pp. 331–334. DOI: 10.1007/BF02762768.

[717] S. Hod. "Bohr's Correspondence Principle and the Area Spectrum of Quantum Black Holes". *Phys. Rev. Lett.* 81 (20 1998), pp. 4293–4296. DOI: 10.1103/PhysRevLett.81.4293.

[718] J. D. Bekenstein and V. Mukhanov. "Spectroscopy of the quantum black hole". *Phys. Lett.B* 360.1–2 (1995), pp. 7–12. ISSN: 0370-2693. DOI: 10.1016/0370-2693(95)01148-j.

[719] J. D. Bekenstein. *Quantum Black Holes as Atoms*. 1998. arXiv: gr-qc/9710076 [gr-qc].

[720] N. Deppe, L. Heisenberg, H. Inchauspé, L. E. Kidder, D. Maibach, S. Ma, J. Moxon, K. C. Nelli, W. Throwe, and N. L. Vu. *Echoes from Beyond: Detecting Gravitational Wave Quantum Imprints with LISA*. 2024. arXiv: 2411.05645 [gr-qc].

[721] S. Bose, A. Mazumdar, G. W. Morley, H. Ulbricht, M. Toroš, M. Paternostro, A. A. Geraci, P. F. Barker, M. Kim, and G. Milburn. "Spin entanglement witness for quantum gravity". *Phys. Rev. Lett.* 119.24 (2017), p. 240401. DOI: 10.1103/PhysRevLett.119.240401.

[722] C. Marletto and V. Vedral. "Gravitationally induced entanglement between two massive particles is sufficient evidence of quantum effects in gravity". *Phys. Rev. Lett.* 119.24 (2017), p. 240402. DOI: 10.1103/PhysRevLett.119.240402.

[723] S. Bose, A. Mazumdar, M. Schut, and M. Toroš. "Mechanism for the quantum natured gravitons to entangle masses". *Phys. Rev. D* 105.10 (2022), p. 106028. DOI: 10.1103/PhysRevD.105.106028.

[724] G. M. Tino. "Testing gravity with cold atom interferometry: results and prospects". *Quantum Science and Technology* 6.2 (2021), p. 024014. DOI: 10.1088/2058-9565/abd83e.

[725] R. J. Marshman, A. Mazumdar, and S. Bose. "Locality and entanglement in table-top testing of the quantum nature of linearized gravity". *Phys. Rev. A* 101.5 (2020), p. 052110. DOI: 10.1103/PhysRevA.101.052110.

[726] D. C. Moore and A. A. Geraci. "Searching for new physics using optically levitated sensors". *Quantum Science and Technology* 6.1 (2021), p. 014008. DOI: 10.1088/2058-9565/abcf8a.

[727] Y. Margalit, O. Dobkowski, Z. Zhou, O. Amit, Y. Japha, S. Moukouri, D. Rohrlich, A. Mazumdar, S. Bose, C. Henkel, et al. "Realization of a complete Stern-Gerlach interferometer: Toward a test of quantum gravity". *Science advances* 7.22 (2021), eabg2879. DOI: 10.1126/sciadv.abg2879.

[728] O. Bengyat, A. Di Biagio, M. Aspelmeyer, and M. Christodoulou. "Gravity-mediated entanglement between oscillators as quantum superposition of geometries". *Phys. Rev. D* 110.5 (2024). ISSN: 2470-0029. DOI: 10.1103/physrevd.110.056046.

[729] G. Tobar, S. K. Manikandan, T. Beitel, and I. Pikovski. "Detecting single gravitons with quantum sensing". *Nat. Commun.* 15.1 (2024), p. 7229. DOI: 10.1038/s41467-024-51420-8.

[730] J. Foo, R. B. Mann, and M. Zych. "Schrödinger's cat for de Sitter spacetime". *Class. Quantum Grav.* 38.11 (2021), p. 115010. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/abf1c4.

[731] J. Foo, R. B. Mann, and M. Zych. "Schrödinger's black hole cat". *Int. J. Modern Phys. D* 31.14 (2022), p. 2242016. ISSN: 1793-6594. DOI: 10.1142/s0218271822420160.

[732] A.-C. de la Hamette, V. Kabel, E. Castro-Ruiz, and Č. Brukner. "Quantum reference frames for an indefinite metric". *Commun. Phys.* 6.1 (), p. 231. ISSN: 2399-3650. DOI: 10.1038/s42005-023-01344-4.

[733] A.-C. de la Hamette, T. D. Galley, P. A. Höhn, L. Loveridge, and M. P. Müller. *Perspective-neutral approach to quantum frame covariance for general symmetry groups*. 2021. arXiv: 2110.13824 [quant-ph].

[734] C. Cepollaro, A. Akil, P. Cieśliński, A.-C. de la Hamette, and Č. Brukner. *The sum of entanglement and subsystem coherence is invariant under quantum reference frame transformations*. 2024. arXiv: 2406.19448 [quant-ph].

[735] L. Loveridge, T. Miyadera, and P. Busch. "Symmetry, Reference Frames, and Relational Quantities in Quantum Mechanics". *Found. Phys.* 48.2 (2018), pp. 135–198. ISSN: 1572-9516. DOI: 10.1007/s10701-018-0138-3.

[736] T. Carette, J. Głowacki, and L. Loveridge. *Operational Quantum Reference Frame Transformations*. 2024. arXiv: 2303.14002 [quant-ph].

[737] M. Zych, F. Costa, and T. C. Ralph. *Relativity of quantum superpositions*. 2018. arXiv: 1809.04999 [quant-ph].

[738] M. Zych and Č. Brukner. "Quantum formulation of the Einstein equivalence principle". *Nat. Phys.* 14.10 (2018), pp. 1027–1031. DOI: 10.1038/s41567-018-0197-6.

[739] L. Freidel, M. Geiller, and D. Pranzetti. "Edge modes of gravity. Part I. Corner potentials and charges". *J. High Energy Phys.* 2020.11 (2020). ISSN: 1029-8479. DOI: 10.1007/jhep11(2020)026.

[740] W. Donnelly and A. C. Wall. "Entanglement Entropy of Electromagnetic Edge Modes". *Phys. Rev. Lett.* 114 (2015), p. 111603. ISSN: 1079-7114. DOI: 10.1103/physrevlett.114.111603.

[741] W. Donnelly and L. Freidel. "Local subsystems in gauge theory and gravity". *J. High Energy Phys.* 2016.9 (2016). ISSN: 1029-8479. DOI: 10.1007/jhep09(2016)102.

[742] A. J. Speranza. "Local phase space and edge modes for diffeomorphism-invariant theories". *J. High Energy Phys.* 2018.2 (2018). ISSN: 1029-8479. DOI: 10.1007/jhep02(2018)021.

[743] S. Carrozza and P. A. Höhn. "Edge modes as reference frames and boundary actions from post-selection". *J. High Energy Phys.* 2022.2 (2022). ISSN: 1029-8479. DOI: 10.1007/jhep02(2022)172.

[744] H. Gomes, S. Langenscheidt, and D. Oriti. *Boundaries, frames and the issue of physical covariance*. 2025. arXiv: 2412.00993 [physics.hist-ph].

[745] L. Donnay, G. Giribet, H. A. González, and M. Pino. "Extended symmetries at the black hole horizon". *J. High Energy Phys.* 2016 (2016), p. 100. ISSN: 1029-8479. DOI: 10.1007/jhep09(2016)100.

[746] G. Lifschytz and M. Ortiz. "Scalar field quantization on the (2+1)-dimensional black hole background". *Phys. Rev. D* 49.4 (1994), pp. 1929–1943. ISSN: 0556-2821. DOI: 10.1103/physrevd.49.1929.

[747] D. Jennings. "On the response of a particle detector in Anti-de Sitter spacetime". *Class. Quantum Grav.* 27.20 (2010), p. 205005. DOI: 10.1088/0264-9381/27/20/205005.

[748] N. K. Dubey and S. Kolekar. "Harvesting fermionic field entanglement in Schwarzschild spacetime". *Phys. Rev. D* 112 (2 2025), p. 025019. DOI: 10.1103/1mwx-7jmf.

[749] I. J. Membrere, K. Gallock-Yoshimura, L. J. Henderson, and R. B. Mann. "Tripartite Entanglement Extraction from the Black Hole Vacuum". *Advanced Quantum Technologies* 6 (2023), p. 9. ISSN: 2511-9044. DOI: 10.1002/qute.202300125.

[750] S. Barman, I. Chakraborty, and S. Mukherjee. "Entanglement harvesting for different gravitational wave burst profiles with and without memory". *J. High Energy Phys.* 2023 (2023), p. 180. ISSN: 1029-8479. DOI: 10.1007/jhep09(2023)180.

[751] S. Wang, M. R. Preciado–Rivas, and R. B. Mann. *Harvesting Information Across the Horizon*. 2025. arXiv: 2504.00083 [gr-qc].

[752] W. Brenna, R. B. Mann, and E. Martín-Martínez. "Anti-Unruh phenomena". *Phys. Lett. B* 757 (2016), pp. 307–311. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2016.04.002.

[753] L. J. Garay, E. Martín-Martínez, and J. de Ramón. "Thermalization of particle detectors: The Unruh effect and its reverse". *Phys. Rev. D* 94 (10 2016), p. 104048. DOI: 10.1103/PhysRevD.94.104048.

[754] L. de Souza Campos and C. Dappiaggi. "The anti-Hawking effect on a BTZ black hole with Robin boundary conditions". *Phys. Lett. B* 816 (2021), p. 136198. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2021.136198.

[755] E. Martín-Martínez, T. R. Perche, and B. de S. L. Torres. "General relativistic quantum optics: Finite-size particle detector models in curved spacetimes". *Phys. Rev. D* 101 (4 2020), p. 045017. DOI: 10.1103/PhysRevD.101.045017.

[756] F. Giacomini and A. Kempf. "Second-quantized Unruh-DeWitt detectors and their quantum reference frame transformations". *Phys. Rev.D* 105.12 (2022), p. 125001. DOI: 10.1103/physrevd.105.125001.

[757] C. E. Wood and M. Zych. "Quantized mass-energy effects in an Unruh-DeWitt detector". *Phys. Rev.D* 106.2 (2022), p. 025012. DOI: 10.1103/PhysRevD.106.025012.

[758] V. Sudhir, N. Stritzelberger, and A. Kempf. "Unruh effect of detectors with quantized center of mass". *Phys. Rev. D* 103.10 (2021), p. 105023. DOI: 10.1103/PhysRevD.103.105023.

[759] E. P. G. Gale and M. Zych. "Relativistic Unruh–DeWitt detectors with quantized center of mass". *Phys. Rev. D* 107 (5 2023), p. 056023. DOI: 10.1103/PhysRevD.107.056023.

[760] L. C. Barbado, E. Castro-Ruiz, L. Apadula, and Č. Brukner. "Unruh effect for detectors in superposition of accelerations". *Phys. Rev. D* 102.4 (2020), p. 045002.

[761] M. Bañados, C. Teitelboim, and J. Zanelli. "Black hole in three-dimensional spacetime". *Phys. Rev. Lett.* 69.13 (1992), pp. 1849–1851. ISSN: 0031-9007. DOI: 10.1103/physrevlett.69.1849.

[762] S. Carlip. *Quantum Gravity in 2+1 Dimensions*. 2023. arXiv: 2312.12596 [gr-qc].

[763] S. Carlip. *Quantum gravity in 2+1 dimensions*. Vol. 50. Cambridge University Press, 2003. DOI: 10.1017/CBO9780511564192.

[764] S. Deser, R. Jackiw, and G. 't Hooft. "Three-dimensional Einstein gravity: Dynamics of flat space". *Ann. Phys.* 152.1 (1984), pp. 220–235. ISSN: 0003-4916. DOI: 10.1016/0003-4916(84)90085-X.

[765] J. D. Barrow, A. Burd, and D. Lancaster. "Three-dimensional classical spacetimes". *Class. Quantum Grav.* 3.4 (1986), p. 551. DOI: 10.1088/0264-9381/3/4/010.

[766] N. J. Cornish and N. E. Frankel. "Gravitation in 2+1 dimensions". *Phys. Rev. D* 43 (8 1991), pp. 2555–2565. DOI: 10.1103/PhysRevD.43.2555.

[767] H. Bacry and J. Levy-Leblond. "Possible kinematics". *J. Math. Phys.* 9 (1968), pp. 1605–1614. DOI: 10.1063/1.1664490.

[768] G. W. Gibbons and C. E. Patricot. "Newton–Hooke spacetimes, Hpp-waves and the cosmological constant". *Class. Quantum Grav.* 20.23 (2003), pp. 5225–5239. ISSN: 1361-6382. DOI: 10.1088/0264-9381/20/23/016.

[769] P. Bizoń, O. Evnin, and F. Ficek. "A nonrelativistic limit for AdS perturbations". *J. High Energy Phys.* 2018 (2018), p. 113. ISSN: 1029-8479. DOI: 10.1007/jhep12(2018)113.

[770] S. K. Pal. "On the magnetic 2+1- D space-time and its non-relativistic counterpart". *J. Phys. Conf. Ser.* 3152.1 (2025), p. 012028. ISSN: 1742-6596. DOI: 10.1088/1742-6596/3152/1/012028.

[771] S. Carlip. "Conformal field theory, (2 + 1)-dimensional gravity and the BTZ black hole". *Class. Quantum Grav.* 22.12 (2005), R85–R123. ISSN: 1361-6382. DOI: 10.1088/0264-9381/22/12/r01.

[772] P. Kraus. "Lectures on Black Holes and the AdS3/CFT2 Correspondence". *Supersymmetric Mechanics - Vol. 3: Attractors and Black Holes in Supersymmetric Gravity*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 1–55. ISBN: 978-3-540-79523-0. DOI: 10.1007/978-3-540-79523-0_4.

[773] R. M. Wald. "The thermodynamics of black holes". *Living Rev. Rel.* 4 (2001), p. 6. DOI: 10.12942/lrr-2001-6. arXiv: gr-qc/9912119.

[774] J. Santiago and M. Visser. "Tolman temperature gradients in a gravitational field". *Eur. J. Phys.* 40.2 (2019), p. 025604. ISSN: 1361-6404. DOI: 10.1088/1361-6404/aaff1c.

[775] C. A. U. Lima, F. Brito, J. A. Hoyos, and D. A. T. Vanzella. "Probing the Unruh effect with an accelerated extended system". *Nat. Commun.* 10 (2019), p. 3030. ISSN: 2041-1723. DOI: 10.1038/s41467-019-10962-y.

[776] K. Shiraishi and T. Maki. "Quantum fluctuation of stress tensor and black holes in three dimensions". *Phys. Rev. D* 49.10 (1994), pp. 5286–5294. ISSN: 0556-2821. DOI: 10.1103/physrevd.49.5286.

[777] S. Avis, C. Isham, and D. Storey. "Quantum field theory in anti-de Sitter space-time". *Phys. Rev. D* 18.10 (1978), p. 3565. DOI: 10.1103/PhysRevD.18.3565.

[778] F. Bussola, C. Dappiaggi, H. R. C. Ferreira, and I. Khavkine. "Ground state for a massive scalar field in the BTZ spacetime with Robin boundary conditions". *Phys. Rev. D* 96.10 (2017). ISSN: 2470-0029. DOI: 10.1103/physrevd.96.105016.

[779] W. G. Unruh. "Notes on black-hole evaporation". *Phys. Rev. D* 14 (4 1976), pp. 870–892. DOI: 10.1103/PhysRevD.14.870.

[780] L. C. B. Crispino, A. Higuchi, and G. E. A. Matsas. "The Unruh effect and its applications". *Rev. Modern Phys.* 80.3 (2008), pp. 787–838. ISSN: 1539-0756. DOI: 10.1103/revmodphys.80.787.

[781] B. S. DeWitt. *General relativity: an Einstein centenary survey*. Ed. by S. Hawking and W. Israel. Cambridge University Press, 1979, p. 680.

[782] E. Martín-Martínez, M. Montero, and M. del Rey. "Wavepacket detection with the Unruh-DeWitt model". *Phys. Rev. D* 87.6 (2013), p. 064038. DOI: 10.1103/PhysRevD.93.024019.

[783] D. Hümmer, E. Martín-Martínez, and A. Kempf. "Renormalized Unruh-DeWitt particle detector models for boson and fermion fields". *Phys. Rev. D* 93.2 (2016), p. 024019. DOI: 10.1103/physrevd.93.024019.

[784] M. S. Ali, S. Bhattacharya, and K. Lochan. *Unruh-DeWitt detector responses for complex scalar fields in de Sitter spacetime*. 2021. arXiv: 2003.11046 [hep-th].

[785] F. M. Guedes, M. S. Guimaraes, I. Roditi, and S. P. Sorella. "A Study of Spin 1 Unruh–DeWitt Detectors". *Universe* 10.8 (2024), p. 307. ISSN: 2218-1997. DOI: 10.3390/universe10080307.

[786] C. Lima, E. Patterson, E. Tjoa, and R. B. Mann. "Unruh phenomena and thermalization for qudit detectors". *Phys. Rev. D* 108.10 (2023), p. 105020. ISSN: 2470-0029. DOI: 10.1103/physrevd.108.105020.

[787] B. d. S. L. Torres. "Particle detector models from path integrals of localized quantum fields". *Phys. Rev. D* 109 (6 2024), p. 065004. DOI: 10.1103/PhysRevD.109.065004.

[788] J. Foo and M. Zych. *Superpositions of thermalisation states in relativistic quantum field theory*. 2023. arXiv: 2307.02593 [quant-ph].

[789] L.-Q. Chen and F. Giacomini. *Quantum Reference Fields Transformations in Linearized Quantum Gravity*. 2026. arXiv: 2606.09344 [gr-qc].

[790] P. A. Höhn, A. Russo, and A. R. H. Smith. *Matter relative to quantum hypersurfaces*. 2023. arXiv: 2308.12912 [quant-ph].

[791] T. Thiemann. *(Quantum) reference frames, relational observables, gauge reduction and physical interpretation*. 2026. arXiv: 2603.04072 [quant-ph].

[792] S. E. Aguilar-Gutierrez, R. Ferrero, P. A. Höhn, and L. Marchetti. *Relational path integral, effective actions and quantum frame covariance in gravity*. 2026. arXiv: 2607.21463 [hep-th].

[793] L. Loveridge, P. Busch, and T. Miyadera. "Relativity of quantum states and observables". *EPL* 117.4 (2017), p. 40004. DOI: 10.1209/0295-5075/117/40004. arXiv: 1604.02836 [quant-ph].

[794] B. Dittrich. "Partial and complete observables for Hamiltonian constrained systems". *Gen. Relativ. Gravit.* 39.11 (2007), pp. 1891–1927. ISSN: 1572-9532. DOI: 10.1007/s10714-007-0495-2.

[795] J. Tambornino. "Relational Observables in Gravity: a Review". *Symmetry, Integrability and Geometry: Methods and Applications* (2012). ISSN: 1815-0659. DOI: 10.3842/sigma.2012.017.

[796] P. Busch, L. Loveridge, and T. Miyadera. "Approximating relational observables by absolute quantities: a quantum accuracy-size trade-off". *J. Phys. A* 49.18 (2016), p. 185301. DOI: 10.1088/1751-8113/49/18/185301. arXiv: 1510.02063 [quant-ph].

[797] S. D. Bartlett, T. Rudolph, R. W. Spekkens, and P. S. Turner. "Degradation of a quantum reference frame". *New Journal of Physics* 8.4 (2006), pp. 58–58. ISSN: 1367-2630. DOI: 10.1088/1367-2630/8/4/058.

[798] S. D. Bartlett, T. Rudolph, and R. W. Spekkens. "Reference frames, superselection rules, and quantum information". *Reviews of Modern Physics* 79.2 (2007), pp. 555–609. ISSN: 1539-0756. DOI: 10.1103/revmodphys.79.555.

[799] S. Ali Ahmad, T. D. Galley, P. A. Höhn, M. P. E. Lock, and A. R. H. Smith. "Quantum Relativity of Subsystems". *Physical Review Letters* 128.17 (2022). ISSN: 1079-7114. DOI: 10.1103/physrevlett.128.170401.

[800] K. Giesel, V. Kabel, and W. Wieland. *Linking Edge Modes and Geometrical Clocks in Linearized Gravity*. 2025. arXiv: 2410.17339 [gr-qc].

[801] C. J. Fewster, D. W. Janssen, and K. Rejzner. *Semi-local observables, edge modes and quantum reference frames in quantum electromagnetism: an algebraic approach*. 2025. arXiv: 2508.20939 [math-ph].

[802] E. Castro-Ruiz and O. Oreshkov. *Relative subsystems and quantum reference frame transformations*. 2023. arXiv: 2110.13199 [quant-ph].

[803] G. Doat and A. Vanrietvelde. "What can we do in a symmetry-constrained perspective? The importance of the total charge's status in quantum reference frame frameworks". *Quantum* 10 (2026), p. 2126. ISSN: 2521-327X. DOI: 10.22331/q-2026-06-08-2126.

[804] S. L. Braunstein, M. Faizal, L. M. Krauss, F. Marino, and N. A. Shah. "Analogue simulations of quantum gravity with fluids". *Nat. Rev. Phys.* 5 (2023), pp. 612–622. DOI: 10.1038/s42254-023-00630-y.

[805] W. G. Unruh. "Experimental Black-Hole Evaporation?" *Phys. Rev. Lett.* 46 (21 1981), pp. 1351–1353. DOI: 10.1103/PhysRevLett.46.1351.

[806] L. J. Garay, J. R. Anglin, J. I. Cirac, and P. Zoller. "Sonic Analog of Gravitational Black Holes in Bose-Einstein Condensates". *Phys. Rev. Lett.* 85 (22 2000), pp. 4643–4647. DOI: 10.1103/PhysRevLett.85.4643.

[807] S. Weinfurtner, E. W. Tedford, M. C. J. Penrice, W. G. Unruh, and G. A. Lawrence. "Measurement of Stimulated Hawking Emission in an Analogue System". *Phys. Rev. Lett.* 106 (2 2011), p. 021302. DOI: 10.1103/PhysRevLett.106.021302.

[808] C. J. Fewster, B. A. Juárez-Aubry, and J. Louko. "Waiting for Unruh". *Class. Quantum Grav.* 33.16 (2016), p. 165003. ISSN: 1361-6382. DOI: 10.1088/0264-9381/33/16/165003.

[809] A. Ashtekar. *Geometry and Physics of Null Infinity*. 2015. arXiv: 1409.1800 [gr-qc].

[810] A. Ashtekar, M. Campiglia, and A. Laddha. "Null infinity, the BMS group and infrared issues". *Gen. Relativ. Gravit* 50.11 (2018). arXiv:1808.07093 [gr-qc], p. 140. ISSN: 0001-7701, 1572-9532. DOI: 10.1007/s10714-018-2464-3.

[811] K. Prabhu and G. Satishchandran. "Infrared finite scattering theory: scattering states and representations of the BMS group". *J. High Energy Phys.* 2024.8 (2024). ISSN: 1029-8479. DOI: 10.1007/jhep08(2024)055.

[812] T. Thiemann. "Gauge field theory coherent states (GCS): I. General properties". *Class. Quantum Grav.* 18.11 (2001), pp. 2025–2064. ISSN: 1361-6382. DOI: 10.1088/0264-9381/18/11/304.

[813] J. R. Klauder and B.-S. Skagerstam. *Coherent states: applications in physics and mathematical physics*. World Scientific, 1985. DOI: 10.1142/0096.

[814] W.-M. Zhang, D. H. Feng, and R. Gilmore. "Coherent states: Theory and some applications". *Rev. Mod. Phys.* 62 (4 1990), pp. 867–927. DOI: 10.1103/RevModPhys.62.867.

[815] C. Kiefer. *Quantum gravity*. Vol. 124. Oxford: Clarendon, 2004. ISBN: 978-0-19-958520-5.

[816] J. B. Hartle and S. W. Hawking. "Wave function of the Universe". *Phys. Rev. D* 28 (12 1983), pp. 2960–2975. DOI: 10.1103/PhysRevD.28.2960.

[817] K. Bradler. *Unitary evolution of a pair of Unruh–DeWitt detectors calculated efficiently to an arbitrary perturbative order*. 2018. arXiv: 1608.08274 [hep-th].

[818] S.-W. Han, W. Chen, L. Chen, Z. Ouyang, and J. Feng. *Reliable quantum master equation of the Unruh–DeWitt detector*. 2025. arXiv: 2502.06411 [gr-qc].

[819] E. G. Brown, E. Martín-Martínez, N. C. Menicucci, and R. B. Mann. "Detectors for probing relativistic quantum physics beyond perturbation theory". *Phys. Rev. D* 87 (2013), p. 084062. ISSN: 1550-2368. DOI: 10.1103/physrevd.87.084062.

[820] J. Polo-Gómez and E. Martín-Martínez. "Nonperturbative method for particle detectors with continuous interactions". *Phys. Rev. D* 109 (2024), p. 045014. ISSN: 2470-0029. DOI: 10.1103/physrevd.109.045014.

[821] S. Carlip. "Black hole thermodynamics". *Int. J. Modern Phys. D* 23.11 (2014), p. 1430023. ISSN: 1793-6594. DOI: 10.1142/s0218271814300237.

[822] R. Emparan and I. Sachs. "Quantization of AdS3 Black Holes in External Fields: Semiclassical Results from Pure Gravity". *Phys. Rev. Lett.* 81.12 (1998), p. 2408. DOI: 10.1103/PhysRevLett.81.2408.

[823] C. Fronsdal. "Completion and Embedding of the Schwarzschild Solution". *Phys. Rev.* 116 (3 1959), pp. 778–781. DOI: 10.1103/PhysRev.116.778.

[824] J. Russo and P. Townsend. "Accelerating branes and brane temperature". *Class. Quantum Grav.* 25.17 (2008), p. 175017. DOI: 10.1088/0264-9381/25/17/175017.

[825] S. Deser and O. Levin. "Accelerated detectors and temperature in (anti-) de Sitter spaces". *Class. Quantum Grav.* 14.9 (1997), p. L163. DOI: 10.1088/0264-9381/14/9/003.

[826] S. Deser and O. Levin. "Equivalence of Hawking and Unruh temperatures and entropies through flat space embeddings". *Class. Quantum Grav.* 15.12 (1998), p. L85. DOI: 10.1088/0264-9381/15/12/002.

[827] S. Deser and O. Levin. "Mapping Hawking into Unruh thermal properties". *Phys. Rev. D* 59.6 (1999), p. 064004. DOI: 10.1103/PhysRevD.59.064004.

[828] A. Sheykin, D. Solovyev, and S. Paston. "Global embeddings of BTZ and Schwarzschild-AdS type black holes in a flat space". *Symmetry* 11.7 (2019), p. 841. DOI: 10.3390/sym11070841.

[829] A. A. Sheykin, M. V. Markov, and S. A. Paston. "Global embedding of BTZ spacetime using generalized method of symmetric embeddings construction". *J. Math. Phys.* 62.10 (2021), p. 102502. ISSN: 0022-2488. DOI: 10.1063/5.0062060.

[830] S. A. Paston. "When does the Hawking into Unruh mapping for global embeddings work?" *J. High Energy Phys.* 122 (2014), p. 6. ISSN: 1029-8479. DOI: 10.1007/jhep06(2014)122.

[831] J. Maldacena and L. Susskind. "Cool horizons for entangled black holes". *Fortschr. Phys.* 61.9 (2013), pp. 781–811. ISSN: 1521-3978. DOI: 10.1002/prop.201300020.

[832] L. Susskind. *ER=EPR, GHZ, and the Consistency of Quantum Measurements*. 2014. arXiv: 1412.8483 [hep-th].

[833] L. Susskind. "Copenhagen vs Everett, Teleportation, and ER=EPR". *Fortschr. Phys.* 64 (2016), pp. 551–564. ISSN: 0015-8208. DOI: 10.1002/prop.201600036.

[834] Y. Yīng, D. Schmid, and R. W. Spekkens. *A contextual advantage for conclusive exclusion: repurposing the Pusey-Barrett-Rudolph construction*. 2026. arXiv: 2512.04173 [quant-ph].

[835] H. Reichenbach. *The direction of time*. Vol. 65. Univ of California Press, 1991.

[836] H. D. Zeh. *The physical basis of the direction of time*. Vol. 4. Springer, 2007. DOI: 10.1007/978-3-540-38861-6.

[837] S. W. Hawking. "Arrow of time in cosmology". *Phys. Rev. D* 32 (10 1985), pp. 2489–2495. DOI: 10.1103/PhysRevD.32.2489.

[838] G. Chiribella and Z. Liu. "Quantum operations with indefinite time direction". *Communications Physics* 5 (2022), p. 190. DOI: 10.1038/s42005-022-00967-3.

[839] A. J. Parzygnat and J. Fullwood. "From Time-Reversal Symmetry to Quantum Bayes' Rules". *PRX Quantum* 4.2 (2023). ISSN: 2691-3399. DOI: 10.1103/prxquantum.4.020334.

[840] A. Bassi, A. Großardt, and H. Ulbricht. "Gravitational decoherence". *Class. Quantum Grav.* 34.19 (2017), p. 193002. ISSN: 1361-6382. DOI: 10.1088/1361-6382/aa864f.

[841] C. Anastopoulos and B. L. Hu. "A master equation for gravitational decoherence: probing the textures of spacetime". *Class. Quantum Grav.* 30.16 (2013), p. 165007. ISSN: 1361-6382. DOI: 10.1088/0264-9381/30/16/165007.

[842] J. Oppenheim, C. Sparaciari, B. Šoda, and Z. Weller-Davies. "Gravitationally induced decoherence vs space-time diffusion: testing the quantum nature of gravity". *Nat. Commun.* 14.1 (2023). ISSN: 2041-1723. DOI: 10.1038/s41467-023-43348-2.

[843] C. Anastopoulos and B.-L. Hu. "Gravitational decoherence: A thematic overview". *AVS Quantum Sci.* 4.1 (2022). ISSN: 2639-0213. DOI: 10.1116/5.0077536.

[844] L. Petruzziello and F. Illuminati. "Quantum gravitational decoherence from fluctuating minimal length and deformation parameter at the Planck scale". *Nat. Commun.* 12.1 (2021). ISSN: 2041-1723. DOI: 10.1038/s41467-021-24711-7.

[845] G. H. S. Aguiar and G. E. A. Matsas. "Simple gravitational self-decoherence model". *Phys. Rev. D* 112.4 (2025). ISSN: 2470-0029. DOI: 10.1103/thcr-22nb.

[846] I. Pikovski, M. Zych, F. Costa, and Č. Brukner. "Universal decoherence due to gravitational time dilation". *Nat. Phys.* 11.8 (2015), pp. 668–672. ISSN: 1745-2481. DOI: 10.1038/nphys3366.

[847] M. P. Blencowe. "Effective Field Theory Approach to Gravitationally Induced Decoherence". *Phys. Rev. Lett.* 111.2 (2013). ISSN: 1079-7114. DOI: 10.1103/physrevlett.111.021302.

[848] F. Giacomini and Č. Brukner. "Quantum superposition of spacetimes obeys Einstein's equivalence principle". *AVS Quantum Science* 4.1 (2022), p. 015601.

[849] B. S. Kay. "Matter–gravity entanglement entropy and the second law for black holes". *Int. J. Modern Phys. D* 32.14 (2023). ISSN: 1793-6594. DOI: 10.1142/s0218271823420026.

[850] W. J. Zeng, F. Labib, and V. Russo. "Towards violations of Local Friendliness with quantum computers". *Quantum* 9 (2025), p. 1851. ISSN: 2521-327X. DOI: 10.22331/q-2025-09-05-1851.

[851] H. Dourdent. "A Gödelian Hunch from Quantum Theory". *Undecidability, Uncomputability, and Unpredictability*. Springer International Publishing, 2021, pp. 97–113. ISBN: 9783030703547. DOI: 10.1007/978-3-030-70354-7_7.

[852] J. Szangolies. "Epistemic Horizons and the Foundations of Quantum Mechanics". *Found. Phys.* 48.12 (2018), pp. 1669–1697. ISSN: 1572-9516. DOI: 10.1007/s10701-018-0221-9.

[853] T. Franzén. *Gödel's Theorem: An Incomplete Guide to Its Use and Abuse*. 1st ed. A K Peters/CRC Press, 2005. DOI: 10.1201/9780429295225.

[854] P. Berglund, T. Hübsch, D. Mattingly, and D. Minic. "Gravitizing the quantum". *Int. J. Modern Phys. D* 31.14 (2022). ISSN: 1793-6594. DOI: 10.1142/s0218277182242024x.

[855] A. Smith. "Gravitational effects of an infinite straight cosmic string on classical and quantum fields: self-forces and vacuum fluctuations". *Proceedings of the Symposium of the Formation and Evolution of Cosmic Strings*. Cambridge University Press. 1990, p. 263.

[856] M. E. Peskin. *An introduction to quantum field theory*. CRC press, 1995. DOI: 10.1201/9780429503559.

[857] D. Tong. *Quantum Field Theory. University of Cambridge Part III Mathematical Tripos*. 2009. URL: https://www.damtp.cam.ac.uk/user/tong/qft.html.

[858] P. Langlois. *Imprints of Spacetime Topology in the Hawking-Unruh Effect*. 2005. arXiv: gr-qc/0510127 [gr-qc].