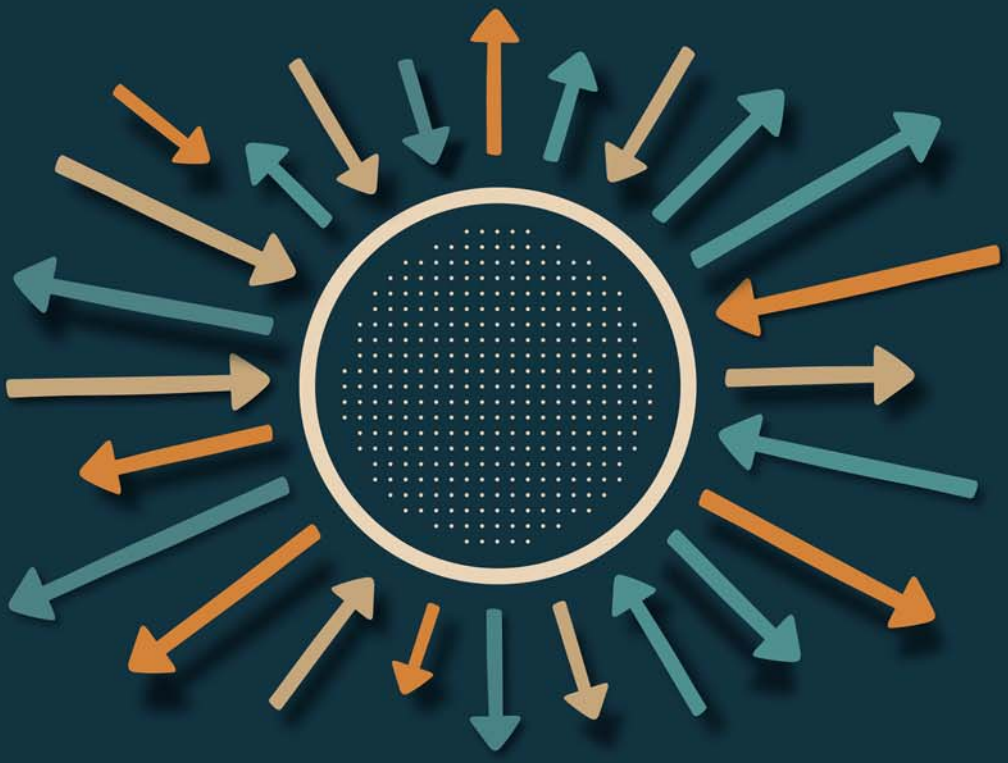
OXFORD

OPEN SYSTEMS

PHYSICS, METAPHYSICS, AND METHODOLOGY



EDITED BY MICHAEL E. CUFFARO
& STEPHAN HARTMANN

Open Systems: Physics, Metaphysics, and Methodology

Open Systems: Physics, Metaphysics, and Methodology

Edited by
MICHAEL E. CUFFARO
STEPHAN HARTMANN

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries.

© Oxford University Press 2026

The moral rights of the authors have been asserted.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, transmitted, used for text and data mining, or used for training artificial intelligence, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by licence or under terms agreed with the appropriate reprographics rights organization. Enquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form
and you must impose this same condition on any acquirer.

Links to third-party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third-party website referenced in this work.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2025950235

ISBN 9780198929246

DOI: 10.1093/9780198929277.001.0001

Printed and bound by
CPI Group (UK) Ltd., Croydon, CR0 4YY

The manufacturer's authorized representative in the EU for product safety is
Oxford University Press España S.A. of Parque Empresarial San Fernando de Henares,
Avenida de Castilla, 2 – 28830 Madrid (www.oup.es/en or product.safety@oup.com).
OUP España S.A. also acts as importer into Spain of products made by the manufacturer.

Contents

List of Figures vii
List of Tables viii
List of Contributors ix
Acknowledgements xi

Introduction 1
Michael E. Cuffaro and Stephan Hartmann

I. THE OPEN SYSTEMS VIEW

1. Quantum Theory is about Open Systems 17
Michael E. Cuffaro and Stephan Hartmann

2. The Temporally Open Systems View 38
Emily Adlam

3. The Relative Nature of Open Quantum Systems 57
Olimpia Lombardi

4. Quantum Systems Other than the Universe 71
David Wallace

5. Frameworks in Physics: Abstractness, Generality, and the Role of Metaphysics 94
Doreen Fraser and Adam Koberinski

II. CONCEPTS OF OPEN AND CLOSED SYSTEMS IN AND BEYOND PHYSICS

6. Newtonian Research and the Open Systems View 113
William L. Harper

7. Blackbody Radiation: The Open and Closed Systems Views and Complementary Reasoning Strategies 129
Molly Kao

8. Boundaries, Frames, and the Issue of Physical Covariance 147
Henrique Gomes, Simon Langenscheidt, and Daniele Oriti

9. Relational Objectivity in the Presence of Finite Quantum Resources 174
Luis C. Barbado and Časlav Brukner

10. Open Systems and Autonomy 193
James Ladyman and Karim P. Y. Thébault

11. Biological Emergence: A Key Exemplar of the Open Systems View 216
George Ellis

III. THE PHYSICS AND METAPHYSICS OF WORLDS AND UNIVERSES

12. Open Systems as Metaphysically Fundamental: Some Questions 233
Jørn Kløvfjell Mjelva, Josh Quirke, and Alastair Wilson

13. How Closed is Cosmology? 247
Sean Gryb and David Sloan

14. Density Matrix Realism 260
Eddy Keming Chen

15. Conservation Laws in the Many-Worlds Interpretation of Quantum Mechanics 283
Lev Vaidman

16. Quantum Theory: Ideals, Infinities, and Pluralities 293
Gemma De les Coves

17. Perspective Duality as a Physical Requirement 317
Wayne C. Myrvold

Index 334

List of Figures

0.1 The action of the sun on the system of Jupiter and its moons 3

1.1 A genealogy of interpretations of standard quantum theory 23

5.1 Some examples of concrete frameworks, theories, and models falling under the abstract field framework 97

9.1 An example of a spin network 184

9.2 An example of a splitting of a one-half spin particle from an end unit *a*, and its later coupling with an end unit *b* 185

9.3 A schematic depiction of a *tentative* form of 'discretized approximate' unitary evolution in time, built up with spin networks 188

List of Tables

11.1 The domain and fundamental principles of the theory of biology 217
11.2 The hierarchical structure of biology in its astronomical context 218

List of Contributors

Emily Adlam is an Assistant Professor in the Department of Philosophy at Chapman University.

Luis C. Barbado is a Postdoctoral Researcher at the Institute for Quantum Optics and Quantum Information Vienna and the Faculty of Physics at the University of Vienna.

Časlav Brukner is a University Professor in the Faculty of Physics at the University of Vienna and Scientific Director of the Institute for Quantum Optics and Quantum Information Vienna.

Eddy Keming Chen is an Associate Professor at the University of California, San Diego, and a Fellow of the John Bell Institute for the Foundations of Physics.

Michael E. Cuffaro is an Alexander von Humboldt Fellow and Visiting Researcher at the Munich Center for Mathematical Philosophy, Ludwig Maximilian University of Munich.

Gemma De les Coves is a Catalan Institution for Research and Advanced Studies (ICREA) Research Professor at the Pompeu Fabra University in Barcelona.

George Ellis is a Fellow of the Royal Society and Emeritus Distinguished Professor in the Department of Mathematics and Applied Mathematics at the University of Cape Town.

Doreen Fraser is a Professor in the Department of Philosophy at the University of Waterloo.

Henrique Gomes is a Junior Research Fellow in Philosophy of Physics at Oriel College, University of Oxford.

Sean Gryb is an Assistant Professor at the University College Groningen of the University of Groningen.

William L. Harper is an Emeritus Professor in the Department of Philosophy at the University of Western Ontario.

Stephan Hartmann is an Alexander von Humboldt Professor and Co-Director of the Munich Center for Mathematical Philosophy, Ludwig Maximilian University of Munich.

Molly Kao is an Associate Professor in the Department of Philosophy at the University of Montreal.

Adam Koberinski is a Postdoctoral Fellow of the Rotman Institute of Philosophy at the University of Western Ontario.

James Ladyman is a Professor in the Department of Philosophy at the University of Bristol.

Simon Langenscheidt is a Ph.D. student at the Arnold Sommerfeld Center for Theoretical Physics, Ludwig Maximilian University of Munich.

Olimpia Lombardi is a Superior Researcher at the University of Buenos Aires and for the Argentine National Scientific and Technical Research Council.

Jørn Kløvfjell Mjelva is a Doctoral Research Fellow in the Department of Philosophy, Classics, History of Art and Ideas at the University of Oslo.

Wayne C. Myrvold is a Professor in the Department of Philosophy at the University of Western Ontario.

Daniele Oriti is a Distinguished Researcher in the Department of Theoretical Physics at Complutense University of Madrid.

Josh Quirke (Ph.D., Birmingham) is an Independent Researcher.

David Sloan is a Reader in Physics at Lancaster University and Chief Scientific Officer of the Foundational Questions Institute.

Karim P. Y. Thébault is an Associate Professor in the Department of Philosophy at the University of Bristol.

Lev Vaidman is an Emeritus Professor in the School of Physics and Astronomy at Tel Aviv University.

David Wallace is the A. W. Mellon Professor in the Department of Philosophy and the Department of History and Philosophy of Science at the University of Pittsburgh.

Alastair Wilson is a Professor in the School of Philosophy, Religion, and History of Science at the University of Leeds, and an Adjunct Senior Research Fellow at Monash University.

Acknowledgements

Thanks to Peter Momtchiloff, Henry Clarke, and Imogene Haslam at Oxford University Press for their support and encouragement at various stages of the publication process, as well as to Christopher von Bülow for creating the index and for providing valuable advice during the copyediting stage. We also want to thank the members of our research group on the philosophy of open systems: James Ladyman, Sébastien Rivat, David Sloan, and Karim Thébault; as well as Florian Boge, John Dougherty, Simon Friederich, Gábor Hofer-Szabó, Nick Huggett, Gregg Jaeger, Alyssa Ney, John Norton, Eric Schliesser, Chris Smeenk, Paul Teller, Roderich Tumulka, Harald Wiltsche, and an anonymous external reader for Oxford University Press for their guidance and advice. Finally, MEC thanks his late friend (and dog), T-Rex, for all the long walks that provided so many opportunities for thinking about the physics of open systems.

This project would not have been possible without the generous financial support of the German Research Council (DFG), through grant number 468374455, and the Alexander von Humboldt Foundation, through an Experienced Researcher Grant to MEC.

Introduction

Michael E. Cuffaro and Stephan Hartmann

The study of open systems is a cornerstone of many scientific disciplines outside physics. Across the biological, social, and cognitive sciences, systems are commonly understood in terms of their interactions with an environment. This is because, at a fundamental level, the entities studied in these disciplines are rarely, if ever, isolated. Organisms exchange energy and matter with their surroundings, social groups depend on external influences, and cognitive processes are shaped by environmental stimuli. Theories that emphasise openness, such as general systems theory (von Bertalanffy 1988), cybernetics (Wiener 1950), synergetics (Haken 1983), and complex systems theory (Ladyman and Wiesner 2020, Thurner et al. 2018), have been instrumental in shaping modern scientific methodologies. These frameworks allow researchers to analyse and predict the behaviour of dynamically interacting components within larger wholes.

While openness is often taken as a given in the special sciences, physics has historically developed with a strong emphasis on closed-system modelling. The success of Newtonian mechanics, classical thermodynamics, and quantum mechanics has to a large extent relied on the ability to isolate and describe systems in terms of well-defined boundary conditions. Even when physical systems are manifestly open, traditional approaches attempt to account for environmental influences by incorporating them into an effectively closed formalism. This methodological preference raises questions with deep philosophical implications: Does this approach adequately capture the complexities of real-world systems, or does it impose artificial constraints on our understanding of nature? To what extent can open-system phenomena be reduced to closed-system descriptions, and what are the limits of this reduction?

Recent developments suggest that the importance of open systems is gaining increasing recognition within physics. Non-equilibrium thermodynamics, pioneered by Prigogine and Nicolis (1977), has challenged the traditional emphasis on closed equilibrium states, highlighting the role of irreversible processes and emergent structures. Similarly, research in quantum biology (Al-Khalili and McFadden 2015) has underscored the interplay between quantum effects and environmental interactions in biological systems. At the intersection of physics and cognitive science, dynamic causal modelling (Weinberger 2020)

Michael E. Cuffaro and Stephan Hartmann, *Introduction*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press. © Oxford University Press (2026).
DOI: 10.1093/9780198929277.003.0001

provides tools for understanding how information flows through complex networks, while interventionist accounts of causation (Pearl 2009, Woodward 2003) emphasise the role of external influences in shaping system behaviour, suggesting that open-system approaches are not merely useful but, in some cases, essential for advancing our understanding of the physical world. In quantum foundations, research into 'quantum causality' has provided new perspectives on the role of environmental interactions in quantum processes. While traditional interpretations of quantum mechanics often rely on a closed-system perspective, recent work in this area challenges this assumption, arguing that a more accurate understanding of the causal structure of quantum phenomena requires that we explicitly consider the openness of the systems involved (Allen et al. 2017, Costa and Shrapnel 2016), shedding new light on the nature of quantum nonlocality and the limits of classical causal explanations (Adlam 2023, Evans 2018, Shrapnel 2019). These developments resonate with broader trends in physics, where the study of complex, adaptive, and information-rich systems increasingly calls for a departure from purely closed-system frameworks.

Despite these advances, the methodological and conceptual challenges posed by open systems remain. How should we formalise openness in physical theories? Are existing mathematical frameworks sufficient, or do we need new theoretical tools? What implications does the shift towards open-system thinking have for the philosophy of science and metaphysics?

In the following chapters, we and our fellow contributors will explore these questions and others in depth. To begin with, however, the next section will provide a short discussion of how open systems are understood and analysed in physics. Along the way, we will introduce key concepts that will be essential as the reader progresses through the book. At the end of this introduction, we will summarise the chapters to come, offering a road map for the journey ahead.

1. Modelling Open Systems in Physics

Physical systems can be closed or open in various senses. A *thermodynamically open* system, for instance, is one that exchanges heat or work with its external environment. Some systems of this kind—consider an internal combustion engine propelling a car down the motorway, for example—inevitably also exchange material particles with their environment. But not all thermodynamically open systems are also *chemically open* in that sense. Any liquid within a container whose walls are thick enough to exclude chemical interactions between the liquid and a bath within which it has been immersed, but still permit the exchange of heat between the liquid and the bath, is *chemically closed* despite being thermodynamically open. No system aside from, perhaps, the entire universe is *absolutely closed*—i.e.,

exchanging nothing at all with its environment—at a minimum, every material system in the universe is subject to the influence of the gravitational field. But in practice, even when a system is not strictly closed we often do successfully model it as such when the influence of the environment on the system can be considered to be negligible, or if there are other principled reasons that allow us to ignore it.

Isaac Newton's Corollary VI to his laws of motion (Newton 1999 [1687]: 423), for instance, states that a system of bodies moving in any way whatsoever with respect to one another will continue to do so in the presence of equal accelerative forces acting on that system along parallel lines. This makes it possible (see Figure 0.1) to treat the system of Jupiter and its moons, for example, as a (quasi-)closed system even though it is being acted upon by the sun. Because of the sun's vast distance from the Jovian system and its relative size, that is, the forces it exerts on that system are approximately equal and parallel, at least up to a certain scale, and can therefore, by Corollary VI, be neglected. For further discussion, see DiSalle (2006), Harper (2011), Harper (this volume), and Saunders (2013).

Unsurprisingly, there are systems—such as lasers, spontaneous emission processes, and (quantum-)information processing systems—for which a proper account requires explicitly considering the environment's influence on the system. In many cases, such a system can be modelled by conceptualising it as enclosed

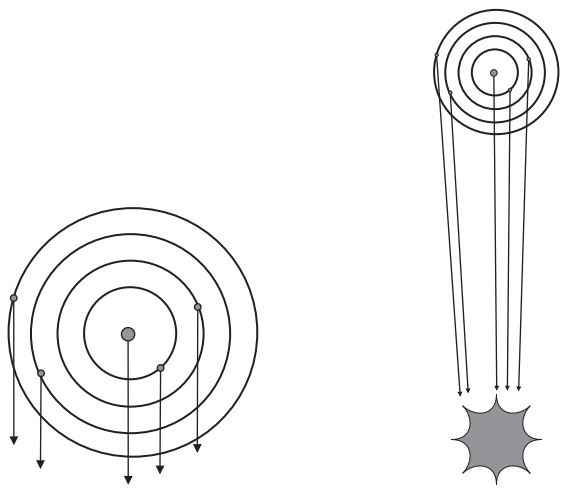


Figure 0.1 The action of the sun on the Jovian system. Although the action of the sun on the Jovian system (the system of Jupiter and its moons) is not exactly equal and parallel (right), because the sun is so far away and so large in comparison to it, its action on the Jovian system is approximately equal and parallel (left), which means that we can appeal to Corollary VI to effectively ignore the action of the sun when reasoning about the Jovian system's internal motions.

within a large 'box', that is itself isolated. That is, we can often accurately represent an (open) system, $\mathcal{S}$, as a subsystem of a larger, isolated composite system, $\mathcal{S} + \mathcal{E}$. This approach is crucial in thermodynamics and statistical mechanics (see Myrvold 2021: chs. 6–8, and Ladyman and Thébault, this volume), as well as in quantum mechanics, which we will now discuss in more detail.

In quantum mechanics, the standard framework (see our own chapter in this volume) relies on wave functions that evolve according to the Schrödinger equation, assuming complete isolation from external influences. However, in reality, most physical systems interact with their environment, leading to more complex behaviour. To address this, physicists use the density matrix formalism, which extends the description of quantum systems to include statistical mixtures of quantum states.

A fundamental distinction in this framework is between pure and mixed states. A *pure state* represents a system with maximal knowledge about its quantum state, meaning it is described by a single wave function. A *mixed state*, by contrast, represents a statistical ensemble of different possible states, reflecting either uncertainty or quantum entanglement (Ismael 2021) arising from the system's interaction with another system. This leads to an important conceptual divide between proper and improper mixtures. *Proper mixtures* arise from classical ignorance—much like flipping a coin and not knowing whether it landed heads or tails. *Improper mixtures*, by contrast, result from abstracting away part of an entangled quantum system (a process known as 'tracing out'). This means that even if a system appears to be in a mixed state, its mixedness may stem from entanglement rather than classical uncertainty.

This distinction is significant for foundational debates in quantum mechanics. If a density matrix describes an improper mixture, should we interpret it as a real physical state of affairs or merely as a tool to describe our partial knowledge? This question relates to the broader discussion of whether quantum states represent reality (i.e., if they are ontic) or knowledge (i.e., if they are epistemic). Furthermore, the density matrix formalism plays a central role in discussions of quantum decoherence (Bacciagaluppi 2025)—the process by which interference effects disappear in macroscopic systems—offering insights into the transition from quantum to classical behaviour.

A key feature of standard quantum mechanics is that the evolution of a system is *unitary*, meaning that information is never lost and probabilities always sum to one. However, when a quantum system interacts with its environment, its evolution is no longer strictly unitary. This leads to dissipation (loss of energy) and decoherence (loss of quantum coherence), which must be accounted for using a more general mathematical framework. The *Lindblad master equation* provides a framework for describing the non-unitary evolution in open quantum systems and ensures that the probabilities here also add up to 1.

The Lindblad equation emerges when we consider a specific small quantum system—such as a two-level atom—and model its interaction with a much larger environment, such as an electromagnetic field or a thermal bath. By 'tracing out' the environmental degrees of freedom, we obtain an effective equation for the system itself, capturing both its unitary evolution (governed by the system's Hamiltonian) and the non-unitary effects of dissipation and decoherence. This derivation relies on certain approximations, such as the Born–Markov assumption, which simplifies the interaction by neglecting memory effects in the environment.

Beyond its practical importance, the Lindblad equation raises fundamental philosophical questions about the nature of quantum theory. If all real-world quantum systems interact with their environment, does this mean that unitary evolution is merely an idealisation? Decoherence, as described by the Lindblad equation, is often invoked to explain why macroscopic systems appear classical—why we do not see Schrödinger's cat in a superposition of alive and dead. But does decoherence truly solve the measurement problem, or does it merely postpone it? Some argue that while decoherence explains why certain quantum states become indistinguishable from classical ones, it does not fully account for why measurements yield definite outcomes. This debate highlights why open quantum systems are not only crucial for practical applications such as quantum computing but also central to a deeper understanding of our most fundamental scientific models, theories, and frameworks.

Understanding the role of open systems in physics requires a careful analysis of how we construct and apply these concepts. In general, a *model* is the key representational tool used in science, and it is used to describe a specific system for a particular epistemic or practical purpose—whether for prediction, explanation, or intervention (Frigg and Hartmann 2020). A *theoretical model*, in turn, is usually formulated in the context of a specific *scientific theory*. When multiple theories share a core set of modeling principles, they form what we call a *theoretical framework*. As we have argued elsewhere (Cuffaro and Hartmann 2024), theoretical frameworks can differ in their underlying *view* of how systems should be conceptualised. For example, the *closed systems view* takes isolated systems as the fundamental objects of study, treating interactions with the environment in terms of couplings between otherwise separable systems that, together, can be regarded as closed. The *open systems view*, by contrast, shifts the focus to the ways in which real systems are inherently embedded in their environments, challenging the assumption that closed-system descriptions are always the most fundamental.

This volume brings together seventeen chapters that examine the physical, metaphysical, and methodological questions surrounding open systems, drawing from a range of disciplines. The contributions explore topics ranging from quantum theory to cosmology, with an interdisciplinary reach that extends into

biology. Organised into three thematic sections, the book investigates: (i) the open systems view in physics, (ii) historical and contemporary perspectives on open- and closed-system concepts, and (iii) the broader, including metaphysical, implications for our understanding of the physical world. By weaving together these different perspectives, this collection demonstrates how shifting the focus to open systems offers fresh insights across fields, reshaping not only our understanding of physical systems but also informing wider debates about the nature of reality.

2. Chapter Summaries

Part I: The Open Systems View

The first part of the book begins with *Quantum Theory is about Open Systems*, wherein we introduce the theoretical framework for quantum physics that is usually taken for granted in philosophical and foundational discussions of the topic: Standard Quantum Theory (ST), and we consider approaches to interpreting the formalism that take ST to be complete in some sense (roughly: '(neo-)Bohrian' and '(neo-)Everettian' approaches). In our chapter, we argue, from the point of view of each, that although ST is formulated in accordance with the closed systems view it is, in fact, ontologically committed to open systems in a way that other theoretical frameworks, which also happen to be formulated in accordance with the closed systems view, are not. We take this to suggest that philosophical and foundational progress in quantum theory may be made, as we have argued in more detail elsewhere, by abandoning what we take to be the artificially restrictive closed systems view in favour of the open systems view.

In Chapter 2, Emily Adlam argues that there is a temporal analogue of the open systems view, which she calls *The Temporally Open Systems View* (TOSV), that suggests that our universe should not be thought of as being divisible into autonomous time slices that are describable independently of one another. Rather, according to TOSV, the happenings at a given time generically depend upon the happenings at other times. Adlam argues that, in fact, quantum histories are not decomposable into autonomous temporal slices in the way required by a temporally closed systems view, and she considers the consequences of TOSV for the ontology of quantum theory and for the question of whether the state of the universe should be thought of as pure or mixed.

The book continues with Chapter 3, by Olimpia Lombardi, entitled *The Relative Nature of Open Quantum Systems*. As Lombardi points out, a quantum system can neither be labelled nor assigned a precise trajectory through spacetime. Moreover, quantum systems generically become correlated with one another in ways that challenge the idea that a composite quantum system is composed of independent

subsystems. As Lombardi explains, this relative nature manifests itself under a number of different forms: the relativity of entanglement, the identification of systems in the context of the process of decoherence, and the many ways in which a composite quantum system can be decomposed into subsystems. She argues that any view, such as our own open systems view, which takes open systems to be fundamental, is required to take subsystem-relativity into account. Indeed, according to Lombardi, neither the open systems view nor the closed systems view should be considered to be fundamental. Rather, they should be thought of as complementary approaches to quantum physics that need not be considered to be rivals.

Chapter 4, *Quantum Systems Other than the Universe*, is by David Wallace. In it, Wallace expounds and develops the claim that physics should be understood as the study of autonomous, not necessarily isolated, dynamical systems, rather than as the project to provide an exact description of the whole universe. In the context of quantum theory this means that, in general, we should take quantum systems to be in mixed states with non-unitary dynamics. In Wallace's framework, however, pure states and unitary dynamics nevertheless continue to have a special role, insofar as the non-unitary dynamics of a system is always taken to be explainable by focusing on a restriction of the dynamics of a larger unitary system to the system of interest, where the larger unitary system comprises both the system of interest and its environment. For Wallace, this follows from what he calls 'four dogmas of quantum reductionism', which he argues for on methodological and metaphysical grounds: (i) 'non-unitarity implies interaction', (ii) 'interaction is derived from dilation', (iii) 'the dilations that explain interaction are to larger unitary systems', and (iv) those dilations are in accordance with a particular class of well-confirmed quantum theories. He then compares this position with our own open systems view.

Frameworks and metaphysical assumptions are the focus for Doreen Fraser and Adam Koberinski, who end this part of the book with Chapter 5, *Frameworks in Physics: Abstractness, Generality, and the Role of Metaphysics*. By contrast with Cuffaro and Hartmann (2024), who associate theoretical frameworks with sets of motivating metaphysical assumptions (such as those associated with the closed systems view in the case of ST), for Fraser and Koberinski, frameworks are, rather, best characterised by their degree of abstractness and their generality of scope. They illustrate the point via detailed analyses of the quantum-field-theoretic and statistical-mechanical frameworks, which, they argue, indicate that metaphysical assumptions are not really integral to the content of a framework, though they may enter at the level of particular theories and models. This makes theoretical frameworks, according to Fraser and Koberinski, poor targets for metaphysical analysis or for relative fundamentality claims.

Part II: Concepts of Open and Closed Systems In and Beyond Physics

In the second part of the book, the focus shifts to analyses of some of the specific ways in which the concepts of open and closed systems are and have been used throughout the history of physics, in contemporary physical theories, and beyond. It begins with Chapter 6, *Newtonian Research and the Open Systems View*, by William L. Harper. Harper contrasts Newton's treatment of the sun's action on Jupiter's moons, with his treatment of its action on our own moon, and argues that this illustrates the importance of what he calls the 'opening-up' of a model; i.e., the incorporation of parameters in one's model of a system through which that model can subsequently be further developed. Harper argues that although Newton's centre-of-mass solution for solar system motions treats our solar system as an isolated system, the extraordinarily successful research programme to develop corrections taking perturbations due to gravitational interactions into account was very much an open systems methodology.

In Chapter 7, *Blackbody Radiation: The Open and Closed Systems Views and Complementary Reasoning Strategies*, Molly Kao recounts the history of the development of Max Planck's blackbody radiation theory, and argues that Planck's treatment of a blackbody system illustrates how both the open and the closed systems view are present in his methodology as two purportedly inconsistent ways of conceptualising the relation between the blackbody and the radiation field. Kao argues that the incorporation of the open systems view, in particular, led to a fundamental, and theoretically fruitful, change in the conception of a physical system. She then considers two different responses to Planck, by James Jeans and Albert Einstein, respectively, and argues that these responses illustrate complementary reasoning strategies that can be thought of as the narrowing and widening of the domain of applicability of a hypothesis, respectively.

Edge modes and boundary charges in gravitational physics, relational dynamics in classical and quantum gravity, and quantum reference frames are the topic of Chapter 8, by Henrique Gomes, Simon Langenscheidt, and Daniele Oriti, entitled *Boundaries, Frames, and the Issue of Physical Covariance*. In their chapter, the authors argue that these research programmes are, in fact, closely linked, all sharing the general aim to achieve a more realistic modelling of gravitational phenomena. Gomes, Langenscheidt, and Oriti explain how removing standardly employed idealisations from gravitational physics makes a model necessarily perspectival, in the sense that it is then always relative to a particular choice of physical reference frame. They close their chapter with a discussion of a number of broader conceptual implications concerning the influence of observers in physics and speculate regarding the physical limits of objectivity.

According to Luis C. Barbado and Časlav Brukner, one way of interpreting the Bell and Kochen–Specker no-go theorems on interpretations of quantum mechanics is as implying that the notions of 'system' and 'experimental context' are fundamentally inseparable, which would further imply that a statement like 'the spin is up in the x-direction' is an inherently relational statement that relates the configuration of a macroscopic device with what one takes to be a property of the system in that context. In this way, the object that textbook quantum mechanics treats as a formally closed system is regarded as not actually physically isolated, since it corresponds to the physical relation the system holds with a macroscopic environment. From the perspective of quantum mechanics, however, referencing a macroscopic device assumes the existence of infinite resources, which, as noted by Barbado and Brukner, leads to conceptual difficulties when applied to situations where such resources are not available. In search of an alternative way of cashing out this approach to the no-go theorems, in Chapter 9, *Relational Objectivity in the Presence of Finite Quantum Resources*, Barbado and Brukner consider the concept of a spin network, originally introduced by Roger Penrose, as a potential way of formalising quantum theory that goes beyond standard quantum mechanics and in particular allows for a description, in the presence of finite resources, that is inherently relational and inseparable between system and experimental context.

The last two chapters of Part II deal with more general questions. According to James Ladyman and Karim P. Y. Thébault, making the distinction between an open and a closed system requires that we, first of all, precisely specify some characteristic quantity that will be conserved if and only if the system is closed. Open systems are then those for which conservation of that quantity fails. This is the core of the conceptual framework, for the disambiguation of scientific language regarding open and closed systems, designed and defended by Ladyman and Thébault in Chapter 10, *Open Systems and Autonomy*. Ladyman and Thébault show how open systems, in this framework, can have well-posed dynamical equations that display various senses of autonomy, and need not always be embeddable within a larger system. They argue that the argument that the universe must be subject to unitary time evolution conflates different senses of closed, and that it is an open question whether we should model our universe as an open or a closed quantum system.

The second part of the book ends with Chapter 11, by George Ellis, entitled *Biological Emergence: A Key Exemplar of the Open Systems View*. According to Ellis, all biological entities are open systems, with various kinds of influences impacting upon them from their surrounding environment, which in turn elicit various kinds of responses. In his chapter, Ellis explains that the existence of modular hierarchical structures is what enables the functional complexity that is required for biological emergence (i.e., the emergence of novel and robust behaviour at a given

level of description) to arise. Ellis then explains that due to an organism's openness to its environment, the precise state of all of its constituent particles is insufficient to determine the future; but that organisms have developed, through Darwinian evolution, processes to cope with this: on short timescales, on developmental and evolutionary timescales, and at the societal level. As for the largest context of all, Ellis asks: 'should the universe itself be regarded as an open system?' Ellis appeals to quantum theory and to cosmology to argue that the answer is yes.

Part III: The Physics and Metaphysics of Worlds and Universes

Cosmology, the universe, and the worlds that, on one interpretation of quantum mechanics, make it up are the subject of the final part of the book, which begins with Chapter 12, *Open Systems as Metaphysically Fundamental: Some Questions*, by Jørn Kløvfjell Mjelva, Josh Quirke, and Alastair Wilson, who consider the question of what to make of the concept of the physical universe, on the supposition that open systems, rather than closed systems, are taken to be metaphysically fundamental. After critically unpacking the idea, expressed in Cuffaro and Hartmann (2024), that the physical universe might be best conceived of as an open system, Mjelva, Quirke, and Wilson then explore and unpack the physical and metaphysical consequences of an alternative precisification of the open systems view, namely the idea that there may simply be no such thing as the cosmos, and that physical reality is indefinitely extensible in the sense that every physical system, without exception, should be thought of as a subsystem of some larger system.

Chapter 13, by Sean Gryb and David Sloan, is entitled *How Closed is Cosmology?* Gryb and Sloan begin with the question: 'What is a closed system?' After critically surveying the standard responses, they propose a solution that leads to a number of remarkable conclusions, such as that the universe is describable as behaving like an open system with a natural arrow of time. They argue that a number of seemingly unusual features of their proposal only seem surprising because of the inadequacy of naive definitions of what it means to be closed. Gryb and Sloan argue that the appropriate notion of closure in cosmology is dynamical closure, and they discuss some further interesting features of models that are closed in this sense; for instance, that in a growing class of models the autonomous system obtained can be integrated through the big bang.

According to Eddy Keming Chen, although being a realist about quantum theory means being a realist about the universal quantum state, realism leaves undecided the question of whether the quantum state of the universe must be pure and represented by a wave function, or whether it can be mixed and represented by a density matrix. The position called *Density Matrix Realism* is an elaboration and defence of the latter position. In Chapter 14, Chen clarifies the thesis by comparing

it with its rival, Wave Function Realism, contrasts it with more general versions of the thesis (including our own open systems view), and addresses a number of what he calls 'frequently asked questions' concerning Density Matrix Realism as it relates to issues such as the intelligibility of the thesis, physical equivalence, and theoretical virtues like Occam's razor.

In Chapter 15, *Conservation Laws in the Many-Worlds Interpretation of Quantum Mechanics*, Lev Vaidman argues that on the many-worlds interpretation of quantum theory, although our universe is represented by a unitarily evolving highly entangled wave function, we only perceive a tiny part of the global wavefunction, and the quantum state of this (i.e., our) 'world' is an effectively open system that evolves non-unitarily. Vaidman then analyses the consequences of questioning the statement that, within a world, conservation laws hold only for ensembles of measurements, on this conception of a world as an open system, and he proposes a possible resolution of the black hole evaporation paradox.

Do closed (quantum) systems actually exist? In Chapter 16, *Quantum Theory: Ideals, Infinities, and Pluralities*, Gemma De les Coves argues that, (i) because no system is ever completely isolated from its environment, (ii) because closure assumes the existence of a radical subject–object split, and (iii) because the description of a closed system necessarily involves an infinity that cannot be instantiated in the world, the answer is no. With regard to (iii), the case of quantum theory is particularly interesting. Quantum theory stands out among other physical theories insofar as describing an ideal closed quantum system does not require an infinity in the dimension of the system. Nevertheless, as De les Coves argues, a number of further infinities are required in the description of such a system even in quantum theory. But if no closed quantum systems exist, how then should we conceive of the universe as a whole? De les Coves puts forth the following thesis: rather than conceiving of the universe as ideal, or singular, it should be conceived of as a plurality.

In the final chapter of the book, Chapter 17, entitled *Perspective Duality as a Physical Requirement*, Wayne C. Myrvold points out that the actual practice of science is unlike the idealised picture that is typically presupposed in much of the philosophy of physics literature, wherein the focus of analysis is on fictitious systems conceived of as wholly isolated from their environments. Myrvold explains how, in practice, one models only a small fraction of a given system's degrees of freedom, and that these constitute, in effect, an open system. Far from being a mere expedient, Myrvold argues that there is a valuable lesson to be learned from this, which is encapsulated in the stance toward physical theory that he calls 'perspective duality': that although (i) a fundamental physical theory should be capable of being formulated without mentioning terms such as 'system', 'apparatus', and 'environment', (ii) any physical theory worth considering as a fundamental theory should also be capable of not embracing everything at once whenever it is used to describe a given system.

Bibliography

Adlam, E. (2023), 'Is There Causation in Fundamental Physics? New Insights from Process Matrices and Quantum Causal Modelling', in *Synthese* 201: 152.

Al-Khalili, J., and McFadden, J. (2015), *Life on the Edge: The Coming of Age of Quantum Biology* (Transworld Publishers).

Allen, J.-M. A., Barrett, J., Horsman, D. C., Lee, C. M., and Spekkens, R. W. (2017), 'Quantum Common Causes and Quantum Causal Models', in *Physical Review X* 7: 031021.

Bacciagaluppi, G. (2025), 'The Role of Decoherence in Quantum Mechanics', in E. N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University), Spring 2025 ed. URL https://plato.stanford.edu/archives/spr2025/entries/qm-decoherence/

Costa, F., and Shrapnel, S. (2016), 'Quantum Causal Modelling', in *New Journal of Physics* 18: 063032.

Cuffaro, M. E., and Hartmann, S. (2024), 'The Open Systems View', in *Philosophy of Physics* 2: 6:1–27.

DiSalle, R. (2006), *Understanding Space-Time* (Cambridge: Cambridge University Press), paperback ed.

Evans, P. W. (2018), 'Quantum Causal Models, Faithfulness, and Retrocausality', in *The British Journal for the Philosophy of Science* 69: 745–774.

Frigg, R., and Hartmann, S. (2020), 'Models in Science', in E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University), Spring 2020 ed. URL https://plato.stanford.edu/archives/spr2020/entries/models-science/

Haken, H. (1983), *Synergetics, an Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology* (New York: Springer).

Harper, W. (2011), *Isaac Newton's Scientific Method: Turning Data Into Evidence* (Oxford: Oxford University Press).

Ismael, J. (2021), 'Quantum Mechanics', in E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University), Fall 2021 ed. URL https://plato.stanford.edu/archives/fall2021/entries/qm/

Ladyman, J., and Wiesner, K. (2020), *What Is a Complex System?* (New Haven: Yale University Press).

Myrvold, W. C. (2021), *Beyond Chance and Credence: A Theory of Hybrid Probabilities* (Oxford: Oxford University Press).

Newton, I. (1999 [1687]), 'Mathematical Principles of Natural Philosophy', trans. I. B. Cohen and A. Whitman, in I. B. Cohen (ed.), *The Principia: A New Translation and Guide*, 371–946 (Berkeley and Los Angeles, California: University of California Press).

Pearl, J. (2009), *Causality: Models, Reasoning, and Inference* (Cambridge: Cambridge University Press), 2nd ed.

Prigogine, I., and Nicolis, G. (1977), *Self-Organization in Non-Equilibrium Systems* (New York: Wiley).

Saunders, S. (2013), 'Rethinking Newton's *Principia*', in *Philosophy of Science* 80: 22–48.

Shrapnel, S. (2019), 'Discovering Quantum Causal Models', in *The British Journal for the Philosophy of Science* 70: 1–25.

Thurner, S., Klimek, P., and Hanel, R. (2018), *Introduction to the Theory of Complex Systems* (Oxford: Oxford University Press).

von Bertalanffy, L. (1988), *General System Theory: Foundations, Development, Applications* (New York: George Braziller).

Weinberger, N. (2020), 'Near-Decomposability and the Timescale Relativity of Causal Representations', in *Philosophy of Science* 87: 841–856.
Wiener, N. (1950), *Cybernetics and Society: The Human Use of Human Beings* (Boston: Houghton Mifflin).
Woodward, J. (2003), *Making Things Happen* (Oxford: Oxford University Press).

I

THE OPEN SYSTEMS VIEW

1
Quantum Theory is about Open Systems

Michael E. Cuffaro and Stephan Hartmann

1.1 Introduction

Most philosophical and foundational discussions of quantum mechanics tend to focus their attention on the unitarily evolving vector, $|\psi\rangle$—the so-called 'wavefunction'—that one takes to be representative of the state of an isolated system at a given time (see, for instance, Harrigan and Spekkens 2010; Ney 2021).[1] Our aim in this chapter is to try to convince anyone who is not already convinced that there are good reasons to focus elsewhere.

To begin with, according to Gleason's theorem (1957), quantum mechanics' assignment of probabilities to the results of experiments on a system is complete, in the sense that every probability measure over yes-or-no questions concerning the system's observable properties is representable, in the most general case, not by a state vector, but by a density operator acting on the system's state space.[2] Insofar as quantum mechanics is an irreducibly probabilistic theoretical framework (in the sense that probabilities are what its predictions consist in), it would seem, therefore, that there is at least a *prima facie* case to be made for focusing on the density operator, rather than on the state vector, in foundational and philosophical discussions.[3]

Besides allowing us to express that a given system has been prepared in one of a number of states, $\{|\psi_i\rangle\}$, with probabilities $\{p_i\}$, a density operator like

$$\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i|, \tag{1}$$

where $|\psi_i\rangle \langle\psi_i|$ is the projection operator associated with the state vector $|\psi_i\rangle$, is also used in quantum mechanics to represent the state of an *open system*, i.e., one that we take to be evolving under the influence of its environment. In this case a density operator like the one given in Eq. (1) is said to represent a so-called 'improper' mixture (D'Espagnat 1966, 1971) of the 'pure states', $\{|\psi_i\rangle\}$—improper because, owing to the fact that the system and its environment are entangled, it is actually impossible to interpret Eq. (1) as literally describing a system that is

Michael E. Cuffaro and Stephan Hartmann, *Quantum Theory is about Open Systems*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press.
© Oxford University Press (2026). DOI: 10.1093/9780198929277.003.0002

in a given pure state $|\psi_i\rangle$ with a given probability p_i (because a subsystem of an entangled system can never be in a pure state).[4]

Of course, when the system and its environment are *not* entangled, we *can* interpret Eq. (1) as representing our ignorance regarding the actual state, $|\psi_i\rangle$, that the system is in. In that case we say that it is in a 'proper' mixture of the pure states $\{|\psi_i\rangle\}$. However we should still take such a statement with a grain of salt, because for a given ensemble whose state is represented by some density operator ρ, there are in general infinitely many preparation procedures that will give rise to it; i.e., $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j| = \sum_k p_k |\phi_k\rangle\langle\phi_k|$, whenever $\sum_j p_j |\psi_j\rangle\langle\psi_j|$ and $\sum_k p_k |\phi_k\rangle\langle\phi_k|$ are related by a unitary transformation (Nielsen and Chuang 2000, 103). And yet, regardless of how a given density operator is decomposed into pure states, its predictions for the statistics of measurements on the system will always be the same.

All of this is according to what we will be calling *standard quantum theory* (ST): a general theoretical framework under which various specific non-relativistic and relativistic quantum theories can be (and have been) formulated.[5] ST is the form of quantum theory that one typically finds in standard textbooks on the topic. In it, the state of a system, $\mathcal{S}$, at a given time t, is represented by a normalised state vector, $|\psi(t)\rangle$, one element of the Hilbert space representing $\mathcal{S}$'s possible states. $\mathcal{S}$'s dynamical evolution is given, in general, by

$$|\psi(t)\rangle = U(t)\,|\psi(0)\rangle, \tag{2}$$

where the time evolution operator $U(t) = \exp(-\mathrm{i}\,H\,t)$, with H a Hermitian operator representing the system's Hamiltonian whose eigenvalues are the possible energy values of the system.

It follows from these assumptions that $U(t)$ is a unitary map on the state space of $\mathcal{S}$. Further, since the system's Hamiltonian does not include terms representing the interaction of the system of interest with another system, it follows that all phenomena must ultimately be modelled,[6] in ST, in terms of unitarily evolving systems represented as isolated from their environment, or what we will be calling *closed systems*. This does not mean, of course, that one cannot model so-called 'open systems phenomena' in ST, but it does mean that the, in general non-unitary, evolution of a density operator representing an open system,[7] is always understood to be merely an effective description of the dynamics of the system. A fundamental description of the dynamics of $\mathcal{S}$, when this is non-unitary, is only obtainable if we consider the dynamics of a larger closed system of which $\mathcal{S}$ is just one part, at any rate according to ST.

In Cuffaro and Hartmann (2024) we introduced an alternative to ST, which we called the *general quantum theory of open systems* (GT), that is interestingly different with regard to (1) the metaphysical position, regarding the fundamental nature of the systems one is modelling, that is used to motivate the adoption of that framework in the first place, as well as in regards to (2) that framework's associated

methodological presuppositions, that constrain how one actually models a system when using the framework. We called the conjunction of (1) and (2) a *view*, and defined the view in accordance with which ST is formulated, namely the *closed systems view*, as:

1. Motivated by the metaphysical position according to which isolated systems are thought of (a priori) as exclusively making up the subject matter of a given scientific domain;
2. Associated with the methodology that models all phenomena in terms of isolated systems; with the consequence that the dynamics of a given open system, $\mathcal{S}$, must always be modelled (whenever the influence of the environment cannot be neglected altogether) in terms of its being coupled to a further system $\mathcal{E}$ (representing its environment), such that $\mathcal{S} + \mathcal{E}$ form an isolated system.

GT, by contrast, is formulated in accordance with the *open systems view* (for a definition, see Cuffaro and Hartmann 2024: 6), the upshot of which is that one does not require the dynamics of an open system to be given, fundamentally, in terms of a larger dynamically isolated system. Rather, in GT, one represents the influence of the environment on $\mathcal{S}$, in general, in the dynamical equations that one takes to govern *its* evolution (i.e., one takes the fundamental dynamical equations for the system to be defined with respect to $\mathcal{S}$, not necessarily with respect to $\mathcal{S} + \mathcal{E}$).

In the remainder of this chapter we will be restricting our focus to ST. In particular, in Cuffaro and Hartmann (2024: 3–4) we argued that despite its many successes, there are a number of *prima facie* reasons to worry whether ST really provides us with all of the conceptual resources we need when it comes to new physics. Some of the issues we mentioned were theoretical in nature, namely the significance of the scale factor in cosmology, the physics of black holes, and quantum reference frames. Two were more general. The first of these, which is epistemic in nature, is that the empirical success of ST ultimately rests upon its applications to open systems, i.e., the subsystems of the universe that we actually have empirical access to, not on its applications to closed systems. Despite the fact that in many cases of interest it is nevertheless possible to effectively treat a given system, $\mathcal{S}$, as though it were isolated, and despite the fact that (as we explain in Cuffaro and Hartmann 2024) even when one cannot do this, one can in many cases model the dynamics of $\mathcal{S}$ by first modeling the dynamics of a larger dynamically coupled system, $\mathcal{S} + \mathcal{E}$, which one can treat as effectively isolated, and then abstracting away from $\mathcal{E}$'s degrees of freedom, at the end of the day it is the dynamics of $\mathcal{S}$, not that of $\mathcal{S} + \mathcal{E}$ (which will typically be highly idealised), that has been thereby confirmed.[8]

The second of the more general reasons we mentioned in Cuffaro and Hartmann (2024) is the one we will be elaborating on in the rest of this chapter. In a nutshell, there are essentially two senses in which one can take any theory of

physics to be complete. On the first, one takes 'complete' to mean that such a theory provides us with, at least in principle, a complete description of reality. On the second, more pragmatic, sense of 'completeness' one takes it to mean that the theory provides us with all of the *resources* we need to describe any (in general probabilistic) given physical phenomenon to whatever level of detail we would like. Approaches to interpreting quantum theory that aim to show that the formalism is complete in the first sense principally include Hugh Everett III's, as well as its many modern descendants (Saunders et al. 2010). Approaches that begin from an understanding of quantum theory as complete in the second sense include those of Niels Bohr, Werner Heisenberg, Wolfgang Pauli, John von Neumann and others, and their modern descendants, and are often referred to using the label 'Copenhagen' or 'orthodox' (Faye 2024), although some have argued that they do not constitute a unified view (see Howard 2004). We will argue, from the point of view of each of these two families of approaches to interpretation, that although ST is formulated in accordance with the closed systems view, it is nevertheless ontologically committed to open systems in a way that other, for instance, classical, theoretical frameworks are not. We will then close by suggesting that philosophical and foundational progress in quantum theory may be made if we adopt an alternative theoretical framework, which might in principle involve a change in corresponding view, within which one can make sense of open systems dynamics in fundamental terms.

1.2 ST is about Open Systems

It is instructive to compare the concept of a state in ST with the corresponding concept in classical theory. In classical theory, completely specifying a system's dynamical state yields a simultaneous answer to all of the experimental questions that one can ask about a given observable quantity associated with that system, questions like: *Is the value of the quantity, A, within the range* Δ*?* In classical theory, once a state assignment is made, the answers to all such questions are determined in advance and irrespective of whether one actually performs an experiment on the system at all. This is, moreover, true for *all* observables. That is, once we assign a state to a system, any question concerning *any* of the observables associated with it is determined in advance of carrying out whatever process allows one to ascertain the answer (cf. Hughes 1989: 61). The same is not true of quantum state descriptions, which suffer from what have been described elsewhere as the 'big' and the 'small' measurement problems.[9]

The big problem (from now on we will omit the inverted commas) is that a given state specification yields, in general, only the probability that the answer to a given experimental question will take on one or another value when asked. This fact—that ST describes phenomena in an irreducibly indeterministic way—is in

itself arguably only a relatively minor departure from classical theory, however, since conditional upon the selection of an observable to measure, one can, in ST, describe the observed probabilities (over the values of that observable) as stemming from an imagined prior classical probability distribution over a corresponding dynamical property of the system that *is* determined in advance by the quantum state (analogously to a statistical-mechanical classical state).

This, of course, is just the flip side of the point, which we made earlier, that there is no way to distinguish, in the context of a local measurement, a properly from an improperly mixed ensemble. That is, conditional upon the selection of a particular observable for measurement, the statistics predicted by the reduced density operator characterising the state of an improperly mixed ensemble will be indistinguishable from the statistics predicted by the density operator characterising an ensemble that is a proper mixture of systems in eigenstates of the selected observable. This means that, conditional upon selecting that particular observable for measurement, we can use this properly mixed state to characterise the system from that point forward, despite the fact that the system has actually become entangled with the measuring apparatus as a result of its interaction with it. The conditional probabilities associated with the various values of the selected observable are given by the Born rule. However, note that the theoretical framework of ST does not include any law for the dynamical evolution of a system described by a state vector other than the Schrödinger equation (or its relativistic analogues). The dynamics that leads to the appearance of one definite outcome of a measurement, rather than another, is not further described by the theory.[10]

This brings us to the small measurement problem, viz., that the various probability distributions that can, conditional upon a corresponding physical interaction, be used to characterise the various observables associated with a system, cannot be embedded into a global prior probability distribution over the values of all observable quantities as they can be for a classically describable system. In quantum theory one can only say that *conditional* upon our inquiring about the observable A, there will be a particular probability distribution that can be used to characterise the possible answers to that question. Conditional upon the selection of a different observable, we will need to use a different probability distribution (over the values of *that* observable) that is in general incompatible with the first.[11] This is usually seen as a problem because ST's unitary description of a measurement interaction does not, by itself, give us an answer to which of these probability distributions we should prefer.[12]

So far nothing we have asserted should be seen as controversial.[13] By contrast, consider the following statements: Because, in classical theory, merely specifying a system's dynamical state (which may or may not be probabilistic as, for instance, in classical statistical mechanics) suffices to fix the answers (or the probabilities over possible answers in the case of a probabilistic state assignment) to all of the possible questions that one can ask concerning the value of any observable associated

with the system, independently of whether a physical interaction through which one can assess that value has actually taken place, we are invited to think of it as originating in the properties of an underlying physical system that exists in a particular way irrespective of anything external, even though there is nothing in the concept of a value, *per se*, that forces us to think of values as originating in this way. This is not the case in quantum theory, where the more complex structure of observables related by it does not similarly invite the inference from the values of observable quantities to the properties of an underlying system in that sense. What is exhibited by the state vector associated with a quantum-mechanical system is not a collection of observer-independent properties, but the quantifiable structure of and interdependencies among the possible ways that one can characterise a system in the context of a physical interaction.[14] This, in any case, is according to the orthodox family of interpretations of quantum theory, which we will have more to say about presently.

There is an ongoing debate over the issue of realism regarding the state vector, where being a realist means that one takes it to represent the observer-independent state of a physical system. In the framework of Harrigan and Spekkens (2010), we call an interpretation of the state vector $|\psi\rangle$-*ontic* if it is taken to represent the real state of a physical system in this sense (cf. Primas 1990: 244). More concretely, a $|\psi\rangle$-ontic interpretation takes it that for a given real state of the system, λ, there is a single state vector corresponding to it. If, further, the state vector is taken to *completely* describe the system's real state (i.e., if there is a one–one mapping between λ and $|\psi\rangle$), then in addition to being $|\psi\rangle$-ontic such an interpretation is $|\psi\rangle$-*complete*. The various interpretations in the lineage of that of Hugh Everett III (1973, 1957), for instance, are generally regarded as both $|\psi\rangle$-ontic and $|\psi\rangle$-complete, while Bohmian mechanics is a $|\psi\rangle$-incomplete interpretation despite being $|\psi\rangle$-ontic.

In the framework of Harrigan and Spekkens (2010) one can also describe so-called $|\psi\rangle$-*epistemic* interpretations of the state vector. Such interpretations take the same real state of a system, λ, to in general be compatible with more than one quantum state. Einstein's interpretation (Demopoulos 2022: ch. 4; Howard 1985) of quantum theory is a $|\psi\rangle$-epistemic interpretation in this sense, and so are its modern successors (e.g., Spekkens 2007). Such interpretations are also, clearly, $|\psi\rangle$-incomplete insofar as they take there to be a real state, λ, that is incompletely described by $|\psi\rangle$.

Orthodox interpretations of the state vector are neither $|\psi\rangle$-ontic nor $|\psi\rangle$-epistemic interpretations in the senses just specified. Yemima Ben-Menahem (2017) calls them 'radical epistemic' interpretations in the sense that for orthodox interpretations of quantum theory, we are not to think of a system as having a real—in the sense of an observer-independent—state, λ, at all, for the reasons suggested by our discussion of the big and especially the small measurement problems above; though it is important to note that the reference to an 'observer' is not necessarily to be understood in the thick metaphysical sense; i.e., as some

agent (embodied or otherwise), and exactly how one interprets the concept of an observer depends on the particular interpretation in the orthodox family that one is talking about.[15]

We mentioned the ontological models framework of Harrigan and Spekkens (2010) because it will be familiar to readers familiar with the Pusey–Barrett–Rudolph theorem (Pusey et al. 2012), which proves that what one might call 'non-radical' epistemic interpretations of a particular sort are impossible.[16] As interpretations in the orthodox family of interpretations of quantum theory cannot be formulated in this framework, however, the ontological models framework turns out to be of rather limited use. A more helpful (though informal) way of relating the various interpretations of quantum theory, given in Janas et al. (2022: ch. 1), is as a genealogy, pictured as a phylogenetic tree with two main branches (see Figure 1.1). One of these branches follows, roughly, the lead of Erwin Schrödinger, for whom the big discovery of quantum mechanics was that a wave phenomenon underlies the particle behaviour of matter. The other, historically older, 'orthodox' branch, roughly follows the lead of Niels Bohr and Werner Heisenberg, for whom the big discovery of quantum mechanics was that solving the problems atomic physics was faced with, in the crisis years of the 1920s, required a new framework for representing physical quantities.[17]

Both of these main branches contain further subdivisions. The former divides into those interpretations that take quantum theory to be complete, i.e., the various interpretations in the (neo-)Everettian family of interpretations which we will have more to say about later, and those that do not, such as Einstein's view as well as the various versions of Bohmian mechanics and dynamical collapse theories.

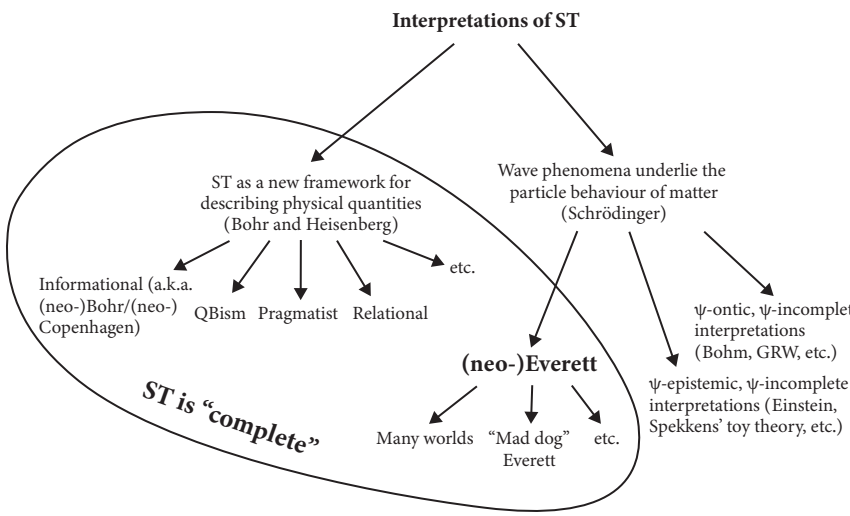


Figure 1.1 A genealogy of interpretations of ST

On the other branch we find the various versions of the so-called 'Copenhagen interpretation' and their descendants. These also take quantum theory to be complete (though in a different sense). They include, but are not limited to, the connected group of ideas of Niels Bohr (1928, 1935, 1949, 1958), Werner Heisenberg (1959, 1971), Grete Hermann (2017b), Wolfgang Pauli (1994), John von Neumann (1955), Carl Friedrich von Weizsäcker (2006), and others. More recent offshoots of this 'orthodox' branch include the objectivist (neo-)Bohrian and (neo-)Copenhagen interpretations of, for instance, Brukner (2017), Bub (2016, 2017), Bub and Pitowsky (2010), Cuffaro (2023), Demopoulos (2012), Demopoulos (2022), Janas et al. (2022), Janas and Janssen (2023), Landsman (2006, 2017), Pitowsky (1989, 2006); the subjectivist interpretation called QBism (Fuchs et al. 2014, Fuchs 2017); Richard Healey's (2012, 2017) pragmatist interpretation; Carlo Rovelli's (1996, 2021) relational interpretation, and others.[18,19]

Orthodox interpretations, which deny that we should think of a system as having a state independently of a given observational context, will truly seem radical for someone who finds it hard to imagine describing reality in an observer-dependent way.[20] Be that as it may, the so-called ideal of the 'detached observer', according to interpreters in the orthodox family, is one that has been found to have a limit to its usefulness (cf. Nagel 1989). Quantum theory, that is, is construed as placing restrictions, not so much on what we can know (although it does that too) but on what we can conceive of knowing (Pitowsky 1994: 113–4; cf. van Fraassen 2002: 3). These restrictions are not absolute (Hermann 2017a; Hermann 2017b: sec. 7; Janas et al. 2022: ch. 7), but they are quantifiable (Bub 2010, 2012; Cuffaro 2020, 2022). One may choose to call such interpretations epistemic if one wishes, but it is clear that this is a very different sense of that word than the one explicated in the ontological models framework of Harrigan and Spekkens.

Despite being 'radically epistemic' (in the above sense), defenders of interpretations in the orthodox family claim that they are realists. However, it is important to be clear on what these interpretations are realist about.[21] What the state vector compactly encodes, on such an interpretation, is a collection of unitarily related conditional probability distributions through which one can characterise a given system in the context of a particular physical interaction.[22] These ways of characterising the system are not themselves (though they may appear to be from within their associated contexts) literal depictions of an observer-independent reality, but they are really constrained, and they are constrained in a particular way that is determined by the quantum state vector independently of any particular observer (d'Espagnat 2001: sec. 4; Healey 2020).

As for the system whose probabilistic characterisations, conditional upon a corresponding physical interaction, are really constrained by the state vector, it is in fact conceived of not as a closed but as an open system on such an interpretation. When a system S is in a state like

$$|\psi_1\rangle = \alpha\,|b_1^+\rangle + \beta\,|b_1^-\rangle, \tag{3}$$

which is unitarily related to the state

$$|\psi_2\rangle = \alpha'\,|b_2^+\rangle + \beta'\,|b_2^-\rangle, \tag{4}$$

and so on, what this signifies is that coupling the degrees of freedom of $\mathcal{S}$ to those of a further system $\mathcal{M}$ will allow one to describe $\mathcal{S}$ in terms of a collection of unitarily related conditional probability distributions over the possible outcomes of an assessment of $\mathcal{M}$ as described with respect to a pointer basis in the Hilbert space of $\mathcal{M}$. Consider Bohr:

> In the treatment of atomic problems, actual calculations are most conveniently carried out with the help of a Schrödinger state function, from which the statistical laws governing observations obtainable under specified conditions can be deduced by definite mathematical operations. It must be recognized, however, that we are here dealing with a purely symbolic procedure, *the unambiguous physical interpretation of which in the last resort requires a reference to a complete experimental arrangement.* Disregard of this point has sometimes led to confusion, and in particular the use of phrases like 'disturbance of phenomena by observation' or 'creation of physical attributes of objects by measurements' is hardly compatible with common language and practical definition. (Bohr 1958: 392–3, our emphasis)

Consider, also, von Neumann:

> [O]ur knowledge of a system $\mathcal{S}'$, i.e., of the structure of a statistical ensemble $\{\mathcal{S}'_1, \mathcal{S}'_2, \ldots\}$, is never described by the specification of a state—or even by the corresponding [vector $|\phi\rangle$]; but usually by the result of measurements performed on the system. (von Neumann 1927: 260)[23]

Consider, finally, an ensemble of identically prepared entangled pairs; i.e., a number of systems, each of which is composed of two interacting subsystems, $\mathcal{S}$ and $\mathcal{M}$, that are entangled together in the state $|\Psi\rangle_{\mathcal{S}+\mathcal{M}}$. If, as we saw earlier, we focus our attention on the sub-ensemble of this ensemble that is formed by choosing $\mathcal{M}$ from each pair, and inquire of the systems in that sub-ensemble concerning the observable O (i.e., if we take a reading of the measurement instrument), we will find that the resulting statistical distribution of answers will be characterisable as having arisen from experiments on a properly mixed ensemble of systems, each of which is in an eigenstate of the selected observable. Strictly speaking, however, in the context of its physical interaction (with $\mathcal{S}_i$) each $\mathcal{M}_i$ *cannot possibly be* in some state $|\phi_j\rangle$ (with probability p_j). Strictly speaking, $\mathcal{S}_i$ and $\mathcal{M}_i$ are open systems

which we describe using reduced density operators (Breuer and Petruccione 2007: viii).

There is no question of taking either the state vector or its generalisation, the density operator, to truly represent a closed system of observer-independent properties on this understanding of quantum theory. The closed system described by $|\psi\rangle_S$ is merely abstract. The real system represented is an open system, most generally described by a density operator. But since ST is formulated in accordance with the closed systems view, conceiving of $\mathcal{S}$ as an open system in this framework requires the introduction of a larger Hilbert space that includes the degrees of freedom of a further system $\mathcal{M}$ that one posits to be coupled with $\mathcal{S}$. This is true regardless of scale, even when $\mathcal{S}$ is the entire universe, assuming it makes sense to model a measurement of the universe at all on one's version of the orthodox interpretation.[24] What a state vector describing its state signifies is that, in the context of an interaction between $\mathcal{S}$ and $\mathcal{M}$ (which could, for instance, be conceived of as the measurement device of some hypothetical being initially not interacting with the system), the outcomes of that interaction would be characterisable in accordance with a specific set of unitarily related conditional probability distributions. Open systems, most generally described by density operators, are the stuff of the world on orthodox interpretations of ST and there is no place in ST's ontology, under such interpretations, for closed systems.[25]

Open systems (in addition to closed systems) are also the stuff of the world for (neo-)Everettian interpretations, which comprise the other main family of interpretations of ST whose defenders, like the defenders of orthodox interpretations, take it to be complete. Note, however, that unlike orthodox interpretations, which begin from an understanding of a complete theory as one that provides us with the resources we need to describe any given (in general, probabilistic) physical phenomenon to whatever level of detail one would like,[26] (neo-)Everettians, consistently with the underlying assumptions of Harrigan & Spekkens' ontological models framework, take a complete physical theory to be one that provides us with, at least in principle, a complete description of physical reality (cf. Everett 1957: 455).

Recall that earlier we discussed the so-called big measurement problem; i.e., the problem raised by the fact that a given state specification yields, in general, only the probability that the answer to a given experimental question will take on one or another value when asked. An advocate of a (neo-)Everettian interpretation solves this problem by asserting that all of the possible outcomes actually obtain in some sense. This does not solve the small measurement problem, however, which is known to (neo-)Everettians as the preferred basis problem. The problem, from the (neo-)Everettian point of view, is that a given state vector can be decomposed in more than one eigenbasis corresponding to the different observables associated with a system. But depending on the eigenbasis chosen we will have

a different story to tell about the 'worlds' (corresponding to possible outcomes) that are described as existing by the state vector.

For instance a spin-$\frac{1}{2}$ system whose state is expressible as $(|z^+\rangle + |z^-\rangle)/\sqrt{2}$ in the eigenbasis of the Pauli Z operator, is expressible in the eigenbasis of the Pauli X operator simply as $|x^+\rangle$. Should we take there to be two worlds in this physical situation, or only one? The problem is solved, at least in part, by appealing to the dynamical process of decoherence. When a system, $\mathcal{S}$, interacts with an external environment—for instance some measurement apparatus, $\mathcal{M}$, which is itself coupled to the rest of the universe outside of the lab—then the terms in the superposition describing $\mathcal{S}$'s state decohere (Zurek 2003) and come to achieve a kind of independence from one another when decomposed into a given eigenbasis of the density operator characterising $\mathcal{M}$. We have seen the effects of this dynamical process already. This is the reason why it is possible to, as we mentioned above, describe a given system $\mathcal{S}_i$, from an improperly mixed ensemble, as if it is in some pure state $|\psi_j\rangle$ with probability p_j even though in reality it cannot be in a pure state since it is entangled with $\mathcal{M}_i$.

How much remains to be solved after taking decoherence into account will depend on the particular interpretation in the (neo-)Everettian family being defended (a point we will return to shortly). But all (neo-)Everettians should agree that the point of view, on one part of the universe, $\mathcal{B}$, of an agent who is herself just one part of the rest of the universe, $\mathcal{A}$, which just happens to be so constituted that it functions as a 'measuring apparatus' (i.e., it can be correlated with different states of $\mathcal{B}$ either by or for the agent), is itself a matter of fact that is completely described by the state vector for the combined interacting system $\mathcal{A} + \mathcal{B}$. In particular, with respect to a given eigenbasis of the density operator characterising $\mathcal{A}$, the state vector for $\mathcal{A} + \mathcal{B}$ predicts statistics for measurements on $\mathcal{B}$ that can be described as stemming from experiments on members of a hypothetical proper mixture of various proportions, p_j, of systems in pure states $|\psi_j\rangle_{\mathcal{B}}$. On a (neo-)Everettian view, each $|\psi_j\rangle_{\mathcal{B}}$ is then taken to represent the world to anyone who is able to conceive of such a notion and who happens to be a part of what is described by the corresponding state, $|\phi_j\rangle_{\mathcal{A}}$, of $\mathcal{A}$. The world, $\mathcal{B}$, as it is represented by an agent who is a part of $\mathcal{A}$, however, is not the world as described by classical physics. The world for the agent is represented by her as a state vector, a conception of the world that diverges from the classical conception of the world in subtle ways that are only manifested when she considers correlated systems, because correlated systems are in general entangled in quantum theory.

This is true not only of a pair of particles, like the ones considered in the EPR experiment (Einstein et al. 1935), that happen to be at hand to be experimented on by the agent in her world. It is also true of $\mathcal{A}$ and $\mathcal{B}$. To say that $\mathcal{A}$ and $\mathcal{B}$ are entangled is, from one way of looking at the situation, to say that all of the various probability distributions over pure states usable by an agent who is a part of $\mathcal{A}$ to describe the statistics she obtains as she interacts with $\mathcal{B}$ in order to assess the

latter's state in the context of various particular measurements, although incompatible with one another if conceived of as all originating in the possessed properties of a single classical object, are nevertheless somehow all consistently encoded in the state vector of the combined system $\mathcal{A} + \mathcal{B}$. This provides some motivation to take them and the worlds (which will, in general, be many) described by each of them to represent something real (cf. Everett 1957: 457),[27] at least for an agent in $\mathcal{A}$ who reflects on the consequences of her being entangled with $\mathcal{B}$.

As we mentioned earlier, where one goes from here will depend on the particular variant of the (neo-)Everettian view being advocated for. A number of questions will be relevant to this. For instance: Should we take *all* of the worlds so describable to (at least potentially) represent something real (Saunders 2021b: sec. 2)? Or should we take only those complex enough and stable enough over the course of their lifetimes to be usefully describable, effectively, as a dynamically evolving classical process (Wallace 2003)? Or should something count as a world only if it is approximately describable using the language of classical physics (in the full sense of that term), and can be recognised as such by a dynamical system that happens to be constituted in the right way to play the role of an agent (Vaidman 2021: sec 2.1)? In fact, not all interpretations in the (neo-)Everettian family are committed to the concept of a world at all (Carroll 2022, Carroll and Singh 2019, Lehner 1997).

Regardless of whether one takes all, some, or even none of the classical conditional probability distributions over pure states encoded in the universal state vector to really represent multiverses, there are very good reasons, we think, from the (neo-)Everettian perspective, to take all such structures to be just as real as anything else described by the universal state vector. In particular, it is a fact about quantum theory—one, indeed, that sets it apart from other possible physical theories (D'Ariano et al. 2017: ch. 1)—that corresponding to any given probability distribution over pure states of $\mathcal{B}$, one can always find a pure state of some larger system $\mathcal{A} + \mathcal{B}$ from which that probability distribution can be derived.[28]

All such 'purifications' of the mixed state of $\mathcal{B}$ are, as D'Ariano et al. (2017: 171) put it, 'essentially the same', in the sense that they are all related to one another by a local unitary transformation on the state of the environment. Thus the density operator, $\rho_{\mathcal{B}}$, decomposed mathematically into a decoherent mixture of various states corresponding to the elements of an eigenbasis of $\rho_{\mathcal{A}}$, is just as objective a description of everything there is, relative to the degrees of freedom included in our representation of $\mathcal{B}$ and to the given eigenbasis, as the universal state vector $|\Psi\rangle_{\mathcal{A}+\mathcal{B}}$. Further, if we would like to describe the mixed state of $\mathcal{B}$ relative to some other basis, we can (equivalently) either unitarily transform $|\Psi\rangle_{\mathcal{A}+\mathcal{B}}$ and then trace over $\mathcal{A}$, or unitarily transform $\rho_{\mathcal{B}}$ directly by applying the Liouville–von Neumann equation. For the (neo-)Everettian there can be no reason that one can glean from the framework of ST *per se* to think of an open system, i.e., a system whose degrees of freedom are coupled with those of its environment, whose state is represented by a density operator $\rho_{\mathcal{B}}$ decomposed in a particular basis—regardless of whether

one's version of (neo-)Everett licenses taking a particular decomposition of the density operator to represent a multitude of worlds or not—as any less an element of reality than the closed system represented by the universal state vector $|\Psi\rangle_{\mathcal{A}+\mathcal{B}}$ (cf. Wallace 2012: secs. 8.4, 8.6; Wallace and Timpson 2010; Vaidman, this volume).

There are no other options for someone who takes ST to be complete, in either the sense of constituting a complete description of physical reality ((neo-)Everettians), or in the sense of providing us with all of the resources we need to describe, to whatever level of detail we would like, any given physical phenomenon (orthodox interpreters). And regardless of whether one favours a (neo-)Everettian or an orthodox interpretation, we have seen that there are good reasons to think of open systems, represented by density operators evolving, in general, non-unitarily, as objectively real by the lights of ST.

1.3 Towards a Wider Framework

ST is formulated in accordance with the closed systems view, which requires that *all* phenomena be modelled in terms of closed systems. And yet, as we have seen in this chapter, the ontology of ST, assuming that one takes it to be complete in some sense, includes much more than just closed systems. ST is about systems that are, in general, open. That the ontological implications of taking a view in a given domain might become decoupled from the metaphysical position that motivates one to adopt that view in the first place might seem paradoxical,[29] but we can begin to understand how this situation arises in quantum theory when we consider that the closed systems view from which ST is formulated is the view of the world that it inherits from classical physics. ST generalises probabilistically (as compared with classical theory) the way in which one represents the state of a closed system while remaining consistent with the methodological presuppositions of the closed systems view. As we have seen, however, the result is that ST is just as ontologically committed to open systems as it is to closed systems, from a (neo-)Everettian perspective, while from an orthodox point of view open systems are the exclusive concern. And yet in all cases ST forces us to model the dynamics of a given open system as merely effective.

Earlier we mentioned that there is an alternative to ST, namely, the theoretical framework that in Cuffaro and Hartmann (2024) we introduced as the general quantum theory of open systems (GT). Unlike ST, according to which the dynamical evolution of a non-unitarily evolving open system is always obtained by first modelling the dynamics of a larger dynamically coupled system, and then abstracting away from the degrees of freedom that one does not care about; in GT, the fundamental dynamical equations that govern the evolution of such a system are defined with respect to that system itself, i.e., they furnish a model of the dynamics

of $\mathcal{S}$, but not necessarily of $\mathcal{S} + \mathcal{E}$ (assuming the existence of $\mathcal{E}$ even makes sense in a given context). Systems are, in this sense, modelled as fundamentally open in general in GT. Moreover this makes a difference: There are dynamical evolutions that one can make physical sense of in GT that one cannot make physical sense of in ST (Cuffaro and Hartmann 2024: sec. 3.3).

1.4 Summary

In the foregoing pages we have argued that unless one demands, pre-theoretically, that every open system be understood as a subsystem of a larger closed system—i.e., unless one takes the closed systems view to have not just methodological force but ontological force as well—then there are good reasons—which come from considering the framework of ST itself—to motivate formulating a more general theoretical framework, either in the form of GT or a framework like it, which might in principle involve a change in corresponding view, within which one can make sense of open systems dynamics in fundamental terms.

Acknowledgements

Thanks to Emily Adlam, Časlav Brukner, Jeff Bub, Chris Fuchs, Richard Healey, Michel Janssen, James Ladyman, Markus Müller, Paul Teller, Wayne Myrvold, and Lev Vaidman for helpful discussions, to two anonymous referees, and to the Press's external reader for their constructive comments on a previous draft.

Endnotes

1. Some important exceptions include Chen (this volume), Maroney (2005, 2008), Robertson (2020), and Weinberg (2014).
2. Gleason's theorem represents measurements as projections and is valid for Hilbert spaces of dimension ≥ 3, though see Busch (2003) for a generalisation. One can always, of course, 'purify' the density operator, ρ_S, representing a system, $\mathcal{S}$, by representing $\mathcal{S}$'s degrees of freedom as coupled with those of a further system, $\mathcal{E}$, such that $\mathcal{S} + \mathcal{E}$'s state is represented by the vector $|\Psi\rangle_{S+\mathcal{E}}$. This will figure importantly later, but for now the relevant point is that $|\Psi\rangle_{S+\mathcal{E}}$ is not defined over the state space, $\mathcal{H}_S$, of $\mathcal{S}$, but over the state space $\mathcal{H}_S \otimes \mathcal{H}_\mathcal{E}$, of $\mathcal{S} + \mathcal{E}$.
3. For two relatively recent philosophical discussions of Gleason's theorem see Acuña (2021) and Demopoulos (2022: ch. 4).
4. In the relativistic setting the situation is even more interesting. In algebraic quantum field theory, or AQFT (for a classic text, see Haag 1996), 'systems' are described in terms of algebras of observables, and 'states' of such systems are defined over those algebras. Classical systems are represented by commutative algebras, whereas non-commutative algebras are used to represent quantum systems. In the case of a non-relativistic quantum system, a local sub-algebra of the global algebra representing a combined system is always a Type I von Neumann factor algebra. In the relativistic setting, the sub-algebras are of Type III. A consequence of this is that one is

unable to prove an analogue of the Kraus representation theorem in AQFT, from which it follows that there is also no analogue of Stinespring's dilation theorem, which is what guarantees that operations defined on the sub-algebras of a system are always extendable to the global algebra (Redei and Valente 2010, 348). Clifton and Halvorson (2001) argue, further, that this precludes one from even being able to define a local operation on a subsystem of an entangled system that could thereby destroy the entanglement between it and the other subsystem, and that quantum field systems are, in this sense, '*unavoidably* and *intrinsically* open to entanglement' (p. 4, emphasis in original). In reply, Valente (2013) argues that Clifton and Halvorson's characterisation of a local operation is too restrictive, and that there is a natural generalisation of the concept that does allow for local disentanglement under certain further conditions.

5. Nielsen and Chuang (2000, 2) explain the difference between a framework and the particular theories associated with it as analogous to the difference between a computer operating system and the particular applications that one can run on it.
6. Note that this is a separate question from that of whether a given system actually *is* isolated.
7. That the system of interest is modelled as a subsystem of an entangled system does not, of course, entail that its dynamics must be non-unitary. From the mere fact that two systems have interacted in the past, it does not follow that they will continue to interact, and in that case the dynamics of the system of interest can be described using a unitary map. This, however, is only a special case of the more general situation that we want to be able to describe, namely the situation in which the environment continues to interact, and in this sense drives the system, over a period of time. To describe this more general scenario one requires a non-unitary map.
8. This is our gloss on an argument suggested to us by Wayne Myrvold (personal communication).
9. The distinction between a 'big' and a 'small' measurement problem, two conceptually distinct aspects of the problem of accounting for the dynamics of measurement in ST (Myrvold 2022: 6), was first introduced (with irony, as implied by the inverted commas) by Itamar Pitowsky (2006), and is further developed in Bub and Pitowsky (2010), Bub (2016), and in Janas et al. (2022). Here we follow the exposition given in Janas et al. (2022).
10. This is a feature of quantum theory that is not interpretation-dependent. As we will discuss in more detail shortly, there are two families of interpretations of ST that take it to be complete in some sense: orthodox and (neo-)Everettian interpretations. Neither of them posit anything other than unitary dynamics to characterise the dynamical evolution of a state vector. In the case of the (neo-)Everettian family of interpretations this is well known. For the case of the orthodox family, see, e.g., Bub (2016: 228); d'Espagnat (2001: secs. 3–4); Howard (2004). Other approaches to interpreting ST (dynamical collapse theories, Bohmian mechanics, and so on) are better thought of as alternative theories rather than alternative interpretations.
11. Specifically, they will be incompatible whenever the operators corresponding to the two observables in question do not commute.
12. What we are here calling the small measurement problem is intimately related to (in the sense that it is precisely what is highlighted by) the thought experiment described in Frauchiger and Renner (2018). For some recent commentaries on this and related thought experiments, see Brukner (2018), Bub (2021), DeBrota et al. (2020), Healey (2018), Lazarovici and Hubert (2019).
13. The statements of the small and big measurement problems are only meant to characterise our observational experience in relation to quantum systems; we have said nothing so far about how to interpret this experience in ontological terms. In some (neo-)Everettian interpretations, for instance, probabilities are grounded in 'branch weights' (Saunders 2021a: sec. 3). Other interpretations will have their own ontological stories to tell about how probabilistic phenomena arise. Nevertheless, each of these ways of interpreting quantum theory needs to be able to either explain, or explain away, the significance of the small and big problems in its own terms.
14. This underlying conception of what a system's state represents is, we think, very similar to what Curiel (2014) has argued is the general conception of a classical state. We will come back to this in note 22 below.
15. For instance, in Janas et al. (2022) and Cuffaro (2023) the concept of an observer is understood schematically (cf. Curiel 2020), associated with a particular choice of 'Boolean frame'. QBism takes the other extreme, and incorporates a metaphysically thick conception of an agent into its

worldview (DeBrota et al. 2020: sec. 2.1, Fuchs 2017: sec. 2). On the relational view an observer is any physical system in interaction with the system of interest (Rovelli 1996: 1641). On the pragmatist interpretation, an observer is associated with a physical context of assessment or 'decoherence environment' within which the truth value of a given statement can be assessed (Healey 2021: sec. 3).

16. For two illuminating discussions of the theorem's assumptions, see Ben-Menahem (2017), Leifer (2014). See also Barrett et al. (2014) and Leifer and Maroney (2013) for related no-go theorems.
17. For more on the genesis of quantum mechanics, see Duncan and Janssen (2019).
18. We do not take this list to be exhaustive. Note also that the version of relational quantum mechanics recently defended by Adlam and Rovelli (2023) arguably does not count as an orthodox interpretation in the sense in which we are using that term here (see Adlam 2022: sec. 4.1, and Adlam and Rovelli 2023: sec. 5). This is in contrast to the approach defended in earlier versions and in Rovelli (2024).
19. Given its radical subjectivity, it may seem strange to call QBism an *orthodox* interpretation (and the same might be said, though perhaps for different reasons, for many of the approaches we mentioned), but this is entirely appropriate given our visualisation of interpretations of quantum mechanics as a phylogenetic tree. QBism begins with the (objectivist) ideas of Bohr, Pauli, Wheeler (1983), and others, and then diverges from them (Fuchs 2017), especially as regards the interpretation of probability (which for QBists is subjective). As far as we can tell, however, QBism and all of the other views we call orthodox share the very general features that we lay out in this section.
20. Just how radical it seems may depend on what one means by an 'observer' here (see above).
21. The various members of the orthodox family have been called 'participatory realist' (Cabello 2017; Fuchs 2016). Peter Evans (2020) has called the kind of objectivity attributed to quantum theory on some versions of the orthodox view 'perspectival objectivity'. Even in QBism, although one interprets probabilities in quantum theory (as one does all probabilities) subjectively, the theory is taken to provide us with an objective method for guiding our action (Fuchs 2017: sec. 2.1).
22. Compare this with Erik Curiel's (2014) construal (esp. sec. 3) of the configuration space of an abstract classical system as encoding a description of its kinematically possible interactions with other abstract classical systems. See also note 14 above.
23. We have slightly changed the mathematical notation to be consistent with our own. For further discussion of this passage see Duncan and Janssen (2013: 251–2).
24. Not everyone who defends an orthodox interpretation of ST will accept that it makes sense to imagine either an actual or hypothetical measurement of the universe. For examples of orthodox interpretations that do accept this (or at any rate see nothing conceptually incoherent in the concept of a measurement of the universe's state), see Bub and Pitowsky (2010: 449) and Janas et al. (2022: 216). Whatever one thinks about this issue, it makes no difference to the point we want to make here, which is that systems are interpreted as open on orthodox interpretations whenever it makes sense to model a system at all using ST (cf. Myrvold, this volume).
25. Note that the statement that 'open systems are the stuff of the world' is compatible with the methodological presuppositions of the closed systems view (from which ST is formulated) because, conditional upon the interaction between $\mathcal{S}$ and $\mathcal{M}$, it is always possible to model the dynamics of $\mathcal{S}$ in terms of a probability distribution over the evolutions, each in accordance with a particular Hamiltonian, of a given set of state vectors. Cf. Myrvold (this volume, sec. 1).
26. See, for instance, Brukner (2018), who writes, quoting Asher Peres: 'While quantum theory can in principle describe anything, a quantum description cannot include everything. In every physical situation something must remain *unanalyzed*. This is not a flaw of quantum theory, but a *logical* necessity . . . ' (5, our emphasis).
27. Properly speaking what the (neo-)Everettian takes to be real are the 'branch weights', i.e., the squared amplitudes corresponding to the terms in a superposition like the one shown in Eq. (1); (objective) probabilities are identified with branch weights (Saunders 2021a: sec. 3).
28. For instance, to purify a mixed state, $\rho_{\mathcal{B}} = \sum_i p_i \, |b_i\rangle_{\mathcal{B}} \, {}_{\mathcal{B}}\langle b_i|$, where the $|b_i\rangle_{\mathcal{B}}$ are elements of an orthonormal basis for $\mathcal{B}$, we introduce an ancilla system $\mathcal{A}$ which has the same state space as

$\mathcal{B}$ and the orthonormal basis $\{|a_i\rangle_\mathcal{A}\}$. We then define the pure state $|\Psi\rangle_{\mathcal{A}+\mathcal{B}} = \sum_i \sqrt{p_i}\,|a_i\rangle_\mathcal{A} \otimes |b_i\rangle_\mathcal{B}$. It can be verified that calculating the reduced density operator for $\mathcal{B}$ in relation to this entangled state yields our original density operator $\rho_\mathcal{B}$ (Nielsen and Chuang 2000: sec. 2.5). The sense in which this sets quantum theory apart from other possible physical theories is explicated by D'Ariano et al. (2017: ch. 1), who list six principles: causality, local discriminability, perfect discriminability, ideal compression, atomicity of composition, and purification, that can be used to characterise theories of (syntactic) information. Classical theory (as applied to information) satisfies the first five. Quantum theory is the only such theory that satisfies all six (D'Ariano et al. 2017: pt. IV).

29. For an analogous gloss, in this case on the development of logic in the 19th and early 20th centuries, see Cassirer (2003 [1910]: ch. 1).

Bibliography

Acuña, P. (2021), 'Must Hidden Variables Theories be Contextual? Kochen & Specker Meet von Neumann and Gleason', in *European Journal for Philosophy of Science* 11: 1–30.

Adlam, E. (2022), 'Does Science Need Intersubjectivity? The Problem of Confirmation in Orthodox Interpretations of Quantum Mechanics', in *Synthese* 200: 1–39.

Adlam, E., and Rovelli, C. (2023), 'Information is Physical: Cross-Perspective Links in Relational Quantum Mechanics', in *Philosophy of Physics* 1: 4.

Barrett, J., Cavalcanti, E. G., Lal, R., and Maroney, O. J. E. (2014), 'No ψ-Epistemic Model Can Fully Explain the Indistinguishability of Quantum States', in *Physical Review Letters* 112: 250403.

Ben-Menahem, Y. (2017), 'The PBR Theorem: Whose Side Is it On?', in *Studies in History and Philosophy of Modern Physics* 57: 80–8.

Bohr, N. (1928), 'The Quantum Postulate and the Recent Development of Atomic Theory', in *Nature* 121: 580–90.

Bohr, N. (1935), 'Can Quantum-Mechanical Description of Physical Reality be Considered Complete?', in *Physical Review* 48: 696–702.

Bohr, N. (1949), 'Discussion with Einstein on Epistemological Problems in Atomic Physics', in P. A. Schilpp (ed.), *The Library of Living Philosophers Volume 7. Albert Einstein: Philosopher-Scientist* (La Salle, Illinois: Open Court), 199–241.

Bohr, N. (1958), 'Quantum Physics and Philosophy', in R. Klibansky (ed.), *Philosophy in the Mid-Century: A Survey* (Firenze: La Nuova Italia Editrice), 308–14. Page reference to reprint in: J. Kalckar (ed.), *Niels Bohr: Collected Works*, Vol. 7 (Amsterdam: Elsevier), 385–94.

Breuer, H.-P., and Petruccione, F. (2007), *The Theory of Open Quantum Systems* (Oxford: Oxford University Press).

Brukner, Č. (2017), 'On the Quantum Measurement Problem', in *Quantum [Un]Speakables II* (Springer), 95–117.

Brukner, Č. (2018), 'A No-Go Theorem for Observer-Independent Facts', in *Entropy* 20: 350.

Bub, J. (2010), 'Quantum Computation: Where Does the Speed-Up Come From?', in A. Bokulich, and G. Jaeger (eds.), *Philosophy of Quantum Information and Entanglement* (Cambridge: Cambridge University Press), 231–46.

Bub, J. (2012), 'Why the Tsirelson Bound?', in Y. Ben-Menahem and M. Hemmo (eds.), *Probability in Physics* (Berlin: Springer), 167–85.

Bub, J. (2016), *Bananaworld: Quantum Mechanics for Primates* (Oxford: Oxford University Press), 2nd paperback edition.
Bub, J. (2017), 'Why Bohr was (Mostly) Right'. arXiv:1711.01604.
Bub, J. (2021), 'Understanding the Frauchiger–Renner Argument', in *Foundations of Physics* 51: 36.
Bub, J., and Pitowsky, I. (2010), 'Two Dogmas about Quantum Mechanics', in S. Saunders, J. Barrett, A. Kent, and D. Wallace (eds.), *Many Worlds? Everett, Quantum Theory, and Reality* (Oxford: Oxford University Press), 433–59.
Busch, P. (2003), 'Quantum States and Generalized Observables: A Simple Proof of Gleason's Theorem', in *Physical Review Letters* 91: 120403.
Cabello, A. (2017), 'Interpretations of Quantum Theory: A Map of Madness', in O. Lombardi, S. Fortin, F. Holik, and C. López (eds.), *What is Quantum Information?* (Cambridge: Cambridge University Press), 138–44.
Carroll, S. M. (2022), 'Reality as a Vector in Hilbert space', in V. Allori (ed.), *Quantum Mechanics and Fundamentality: Naturalizing Quantum Theory between Scientific Realism and Ontological Indeterminacy* (Cham: Springer), 211–24.
Carroll, S. M., and Singh, A. (2019), 'Mad-Dog Everettianism: Quantum Mechanics at its Most Minimal', in A. Aguirre, B. Foster, and Z. Merali (eds.), *What is Fundamental?* (Cham: Springer), 95–104.
Cassirer, E. (2003 [1910]), 'Substance and Function', in *Substance and Function [1910] & Einstein's Theory of Relativity [1921]* (Mineola: Dover), trans. W. C. Swabey.
Clifton, R., and Halvorson, H. (2001), 'Entanglement and Open Systems in Algebraic Quantum Field Theory', in *Studies in History and Philosophy of Modern Physics* 32: 1–31.
Crull, E., and Bacciagaluppi, G. (eds.) (2017), *Grete Hermann: Between Physics and Philosophy* (Dordrecht: Springer).
Cuffaro, M. E. (2020), 'Information Causality, the Tsirelson Bound, and the "Being-Thus" of Things', in *Studies in History and Philosophy of Modern Physics* 72: 266–77.
Cuffaro, M. E. (2022), 'The Philosophy of Quantum Computing', in E. Miranda (ed.), *Quantum Computing in the Arts and Humanities* (Dordrecht: Springer), 107–52.
Cuffaro, M. E. (2023), 'The Measurement Problem is a Feature, Not a Bug – Schematising the Observer and the Concept of an Open System on an Informational, or (Neo-) Bohrian, Approach', in *Entropy* 25: 1410. arXiv:2308.16371.
Cuffaro, M. E., and Hartmann, S. (2024), 'The Open Systems View', in *Philosophy of Physics* 2:6.
Curiel, E. (2014), 'Classical Mechanics is Lagrangian; it is Not Hamiltonian', in *The British Journal for Philosophy of Science* 65: 269–321.
Curiel, E. (2020), 'Schematizing the Observer and the Epistemic Content of Theories'. arXiv:1903.02182v3.
D'Ariano, G. M., Chiribella, G., and Perinotti, P. (2017), *Quantum Theory from First Principles: An Informational Approach* (Cambridge: Cambridge University Press). 2019 paperback edition.
DeBrota, J. B., Fuchs, C. A., and Schack, R. (2020), 'Respecting One's Fellow: QBism's Analysis of Wigner's Friend', in *Foundations of Physics* 50: 1859–74.
Demopoulos, W. (2012), 'Generalized Probability Measures and the Framework of Effects', in Y. Ben-Menahem and M. Hemmo (eds.), *Probability in Physics* (Heidelberg: Springer), 201–17.
Demopoulos, W. (2022), *On Theories* (Cambridge: Harvard University Press).
d'Espagnat, B. (1966), 'An Elementary Note About "Mixtures"', in A. de Shalit, H. Feshbach, and L. van Hove (eds.), *Preludes in Theoretical Physics* (Amsterdam: North Holland Wiley), 185–91.

d'Espagnat, B. (1971), *Conceptual Foundations of Quantum Mechanics* (Menlo Park, CA: W. A. Benjamin). Second edition: 1976.

d'Espagnat, B. (2001), 'Reply to K A Kirkpatrick'. arXiv:quant-ph/0111081v1.

Duncan, A., and Janssen, M. (2013), '(Never) Mind Your *p*s and *q*s: Von Neumann versus Jordan on the Foundations of Quantum Theory', in *The European Physical Journal H* 38: 175–259.

Duncan, A., and Janssen, M. (2019), *Constructing Quantum Mechanics, Vol. 1, The Scaffold, 1900–1923* (Oxford: Oxford University Press).

Einstein, A., Podolsky, B., and Rosen, N. (1935), 'Can Quantum-Mechanical Description of Physical Reality be Considered Complete?', in *Physical Review* 47: 777–80.

Evans, P. W. (2020), 'Perspectival Objectivity', in *European Journal for Philosophy of Science* 10: 19:1–21.

Everett, H., III (1957), '"Relative State" Formulation of Quantum Mechanics', in *Reviews of Modern Physics* 29: 454–62.

Everett, H., III (1973), 'The Theory of the Universal Wave Function' [1955], in B. DeWitt and N. Graham (eds.), *The Many-Worlds Interpretation of Quantum Mechanics* (Princeton: Princeton University Press), 3–140.

Faye, J. (2024), 'Copenhagen Interpretation of Quantum Mechanics', in E. N. Zalta and U. Nodelman (eds.), *Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University), Summer 2024 edition. URL https://plato.stanford.edu/archives/sum2024/entries/qm-copenhagen/

Frauchiger, D., and Renner, R. (2018), 'Quantum Theory Cannot Consistently Describe the Use of Itself', in *Nature Communications* 9: 3711.

Fuchs, C. A. (2016), 'On Participatory Realism', in I. T. Durham and D. Rickles (eds.), *Information and Interaction: Eddington, Wheeler, and the Limits of Knowledge* (Berlin: Springer), 113–34.

Fuchs, C. A. (2017), 'Notwithstanding Bohr, the Reasons for QBism', in *Mind and Matter* 15: 245–300.

Fuchs, C. A., Mermin, N. D., and Schack, R. (2014), 'An Introduction to QBism with an Application to the Locality of Quantum Mechanics', in *American Journal of Physics* 82: 749–54.

Gleason, A. M. (1957), 'Measures on the Closed Subspaces of a Hilbert Space', in *Journal of Mathematics and Mechanics* 6: 885–93.

Haag, R. (1996), *Local Quantum Physics—Fields, Particles, Algebras* (Berlin: Springer), 2nd edition.

Harrigan, N., and Spekkens, R. W. (2010), 'Einstein, Incompleteness, and the Epistemic View of Quantum States', in *Foundations of Physics* 40: 125–57.

Healey, R. (2012), 'Quantum Theory: A Pragmatist Approach', in *The British Journal for the Philosophy of Science* 63: 729–71.

Healey, R. (2017), *The Quantum Revolution in Philosophy* (Oxford: Oxford University Press).

Healey, R. (2018), 'Quantum Theory and the Limits of Objectivity', in *Foundations of Physics* 48: 1568–89.

Healey, R. (2020), 'A Pragmatist View of the Metaphysics of Entanglement', in *Synthese* 197: 4265–302.

Healey, R. (2021), 'Scientific Objectivity and its Limits', in *The British Journal for the Philosophy of Science* 75: 639–661.

Heisenberg, W. (1959), *Physics and Philosophy: The Revolution in Modern Science* (London: George Allen & Unwin Ltd).

Heisenberg, W. (1971), *Physics and Beyond* (New York: Harper & Row). Trans. A. J. Pomerans.

Hermann, G. (2017a), 'Determinism and Quantum Mechanics' [1933], trans. E. Crull and G. Bacciagaluppi, in Crull and Bacciagaluppi (2017: 223–38).

Hermann, G. (2017b), 'Natural-Philosophical Foundations of Quantum Mechanics' [1935], trans. E. Crull and G. Bacciagaluppi, in Crull and Bacciagaluppi (2017: 239–78).

Howard, D. (1985), 'Einstein on Locality and Separability', in *Studies in History and Philosophy of Science* 16: 171–201.

Howard, D. (2004), 'Who Invented the "Copenhagen Interpretation"? A Study in Mythology', in *Philosophy of Science* 71: 669–82.

Hughes, R. I. G. (1989), *The Structure and Interpretation of Quantum Mechanics* (Cambridge, MA: Harvard University Press).

Janas, M., Cuffaro, M. E., and Janssen, M. (2022), *Understanding Quantum Raffles: Quantum Mechanics on an Informational Approach: Structure and Interpretation* (Cham: Springer).

Janas, M., and Janssen, M. (2023), 'Broken Arrows: Hardy–Unruh Chains and Quantum Contextuality', in *Entropy* 25: 1568.

Landsman, N. P. (2006), 'When Champions Meet: Rethinking the Bohr–Einstein Debate', in *Studies in History and Philosophy of Modern Physics* 37: 212–42.

Landsman, N. P. (2017), *Foundations of Quantum Theory* (Cham: Springer).

Lazarovici, D., and Hubert, M. (2019), 'How Quantum Mechanics Can Consistently Describe the Use of Itself', in *Scientific Reports* 9: 470.

Lehner, C. (1997), 'What it Feels Like to Be in a Superposition. And Why', in *Synthese* 110: 191–216.

Leifer, M. S. (2014), 'Is the Quantum State Real? An Extended Review of ψ-Ontology Theorems', in *Quanta* 3: 67–155.

Leifer, M. S., and Maroney, O. J. E. (2013), 'Maximally Epistemic Interpretations of the Quantum State and Contextuality', in *Physical Review Letters* 110: 120401.

Maroney, O. J. E. (2005), 'The Density Matrix in the de Broglie–Bohm Approach', in *Foundations of Physics* 35: 493–510.

Maroney, O. J. E. (2008), 'The Physical Basis of the Gibbs–von Neumann Entropy'. arXiv:quant-ph/0701127v2.

Myrvold, W. C. (2022), 'Philosophical Issues Raised by Quantum Theory and its Interpretations', in O. Freire (ed.), *The Oxford Handbook of the History of Quantum Interpretations* (Oxford: Oxford University Press), 53–76.

Nagel, T. (1989), *The View from Nowhere* (Oxford: Oxford University Press).

Ney, A. (2021), *The World in the Wave Function* (Oxford: Oxford University Press).

Nielsen, M. A., and Chuang, I. L. (2000), *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press).

Pauli, W. (1994), *Writings on Physics and Philosophy* (Berlin: Springer), C. P. Enz and K. von Meyenn (eds.).

Pitowsky, I. (1989), *Quantum Probability – Quantum Logic* (Hemsbach: Springer).

Pitowsky, I. (1994), 'George Boole's "Conditions of Possible Experience" and the Quantum Puzzle', in *The British Journal for the Philosophy of Science* 45: 99–125.

Pitowsky, I. (2006), 'Quantum Mechanics as a Theory of Probability', in W. Demopoulos and I. Pitowsky (eds.), *Physical Theory and its Interpretation: Essays in Honor of Jeffrey Bub* (New York: Springer), 213–40.

Primas, H. (1990), 'Mathematical and Philosophical Questions in the Theory of Open and Macroscopic Quantum Systems', in A. I. Miller (ed.), *Sixty-Two Years of Uncertainty* (New York: Plenum Press), 233–57.
Pusey, M. F., Barrett, J., and Rudolph, T. (2012), 'On the Reality of the Quantum State', in *Nature Physics* 8: 475–8.
Rédei, M., and Valente, G. (2010), 'How Local are Local Operations in Local Quantum Field Theory?', in *Studies in History and Philosophy of Modern Physics* 41: 346–353.
Robertson, K. (2020), 'In Search of the Holy Grail: How to Reduce the Second Law of Thermodynamics', in *The British Journal for the Philosophy of Science* 73: 849–1124.
Rovelli, C. (1996), 'Relational Quantum Mechanics', in *International Journal of Theoretical Physics* 35: 1637–78.
Rovelli, C. (2021), *Helgoland: Making Sense of the Quantum Revolution* (New York: Riverhead Books).
Rovelli, C. (2024), 'Can Alice Do Science and Have Friends, in a Relational Quantum World? Solipsism and Relational Quantum Mechanics'. arXiv:2410.20012v1.
Saunders, S. (2021a), 'The Everett Interpretation: Probability'. arXiv:2103.03966.
Saunders, S. (2021b), 'The Everett Interpretation: Structure'. arXiv:2103.01366.
Saunders, S., Barrett, J., Kent, A., and Wallace, D. (eds.) (2010), *Many Worlds? Everett, Quantum Theory and Reality* (Oxford: Oxford University Press).
Spekkens, R. W. (2007), 'Evidence for the Epistemic View of Quantum States: A Toy Theory', *Physical Review A* 75: 032110.
Vaidman, L. (2021), 'Many-Worlds Interpretation of Quantum Mechanics', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University), Fall 2018 edition. URL https://plato.stanford.edu/archives/fall2021/entries/qm-manyworlds/
Valente, G. (2013), 'Local Disentanglement in Relativistic Quantum Field Theory', in *Studies in History and Philosophy of Modern Physics* 44: 424–432.
van Fraassen, B. C. (2002), *The Empirical Stance* (New Haven: Yale University Press).
von Neumann, J. (1927), Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik, in *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 245–72.
von Neumann, J. (1955), *Mathematical Foundations of Quantum Mechanics [1932]* (Princeton: Princeton University Press). Trans. R. T. Beyer.
von Weizsäcker, C. F. (2006), *The Structure of Physics* [1985] (Berlin: Springer).
Wallace, D. (2003), 'Everett and Structure', in *Studies in History and Philosophy of Modern Physics* 34: 87–105.
Wallace, D. (2012), *The Emergent Multiverse* (Oxford: Oxford University Press).
Wallace, D., and Timpson, C. G. (2010), 'Quantum Mechanics on Spacetime I: Spacetime State Realism', in *The British Journal for the Philosophy of Science* 61: 697–727.
Weinberg, S. (2014), 'Quantum Mechanics Without State Vectors', in *Physical Review A* 90: 042102.
Wheeler, J. A. (1983), 'Law Without Law', in J. A. Wheeler and W. Zurek (eds.), *Quantum Theory and Measurement* (Princeton: Princeton University Press), 182–213.
Zurek, W. H. (2003), 'Decoherence and the Transition from Quantum to Classical—Revisited'. arXiv:quant-ph/0306072v1.

2
The Temporally Open Systems View

Emily Adlam

As noted presciently by Cuffaro and Hartmann (2024), we sometimes come upon cases in physics where 'the ontological implications of taking a view in a given domain might become decoupled from the metaphysical position that motivates that view in the first place' (Cuffaro and Hartmann 2024: 26). The obvious danger in such cases is that even once we have recognised that the ontological implications of the view are inconsistent with the motivating metaphysical position, we may nonetheless fail to fully extricate ourselves from the original metaphysical picture, which can lead us to end up with a fundamentally confused account of reality. Thus in such circumstances it is important to consciously identify and discard conceptions inherited from metaphysical positions that we no longer adhere to, as Cuffaro and Hartmann set out to do.

Quantum mechanics exemplifies this problem in a number of ways. Prior to quantum mechanics, physics was often conceived of in terms of a metaphysical picture of reality grounded in reductionism and mechanistic accounts of physical reality, and at least some features of the early formulation of quantum mechanics still reflected this, but it has become clear that the ontological implications of quantum mechanics do not fit that picture. We are slowly moving away from that original metaphysical picture and seeking a new one, but the process has been long and complex. Discarding the closed systems view in favour of the open systems view is an important step on that journey; but it is not the final step.

In this spirit, we argue that there is a temporal analogue of the open systems view expounded by Cuffaro and Hartmann (2024). This 'temporally open systems view' (TOSV) says that we should not think of the universe as breaking down into autonomous time-slices whose instantaneous evolution is always independent from events at other times. Rather we should expect that generically happenings at a given time will depend on happenings at other times. Employing the definitions proposed by Cuffaro and Hartmann, we argue that there are good reasons to think that the TOSV should be regarded as ontically fundamental. We use the failure of narratability in quantum mechanics to demonstrate that generically, quantum histories cannot be decomposed into autonomous temporal slices, because that decomposition would fail to include information about global properties relating different slices. We also discuss the status of temporal closure

Emily Adlam, *The Temporally Open Systems View*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press. © Oxford University Press (2026).
DOI: 10.1093/9780198929277.003.0003

in several interpretations of quantum mechanics. Finally, we consider the consequences of the TOSV for the question about whether the state of the universe as a whole is pure.

2.1 The Temporally Open Systems View

It is common for physicists to espouse a form of reductionism which imagines the world as being made up of small autonomous parts (Castellani 2002). Weinberg refers to this vision as 'petty reductionism', which he describes as '(the) doctrine that things behave the way they do because of the properties of their constituents: for instance, a diamond is hard because the carbon atoms of which it is composed can fit together neatly' (Weinberg 2001). We may reasonably assume that what Weinberg means by this is that for an isolated system, the behaviour of that system can be fully accounted for by the behaviour of its parts—the large-distance behaviour completely supervenes on the small-distance behaviour. Butterfield describes a similar view which he calls 'pointillisme as regards space', and defines as 'the doctrine that the instantaneous state of the world is fully described by all the intrinsic properties, at that time, of all spatial points and/or point-sized bits of matter' (Butterfield 2006). The ontological vision of petty reductionism comes along with a methodological approach which suggests that science should proceed by separating systems into their constituent parts, studying the autonomous behaviour of the parts and their modes of interaction with other parts, and thereby learning to deduce the behaviour of the whole from the behaviour of the parts—and although not completely universal, this kind of reductive physicalism is a very standard methodology of science. For example, Weinberg notes that petty reductionism tends to dominate the conversation about reductionism more generally because 'much of the reductive progress in science has been in answering questions about what things are made of' (Weinberg 2001)—that is, much scientific work presupposes and successfully makes use of this specific form of reductionism. We will henceforth refer to both this ontological position and its methodological analogue as 'whole-to-part reductionism'.

In this chapter, we take it that a type of system can be understood as 'autonomous' provided that (a) we can fully characterise the system and its intrinsic properties independently of other systems, and (b) we can specify how the intrinsic properties of the system would evolve independently if it were not in interaction with any other systems. A system may be autonomous even if it is, in actuality, always in interaction with other systems: autonomy simply requires that we can sensibly conceive of and describe the system independently of other systems, and this can be true even if it is not in fact behaving independently in some particular scenario. Whole-to-part reductionism can be understood as positing that nature is made up of fundamental systems which are autonomous in this

sense, i.e., they have independent existence and independent behaviour outside of their interactions.

On the other hand, the open systems view espoused by Cuffaro and Hartmann (2024) tells us that systems are generically *open*, meaning that the influence of the environment on the system cannot be treated as negligible. That is, although we may occasionally encounter closed systems, they should not be regarded as the norm or the theoretic ideal. And systems which are 'open' in this sense are not fully 'autonomous': an open system cannot be fully characterised independently from its environment, and moreover if systems of the relevant kind are *generically* open, there may be little we can meaningfully say about how they would evolve in the counterfactual case where they are completely isolated. This is particularly true for open systems which are entangled with their environment, since in that case many of the physically important properties of the system may be non-local and holistic, so we may be significantly limited in our ability to ascribe intrinsic properties to the system in and of itself. Of course, in such circumstances we can use the Lindblad equation to describe dynamics for the system without direct reference to its environment, but this evolution is derived from the more complete dynamics for a larger composite system, rather than the dynamics of the composite being built up out of the autonomous dynamics of its parts, so at a conceptual level it arguably does not make sense to try to imagine the evolution of the system in a counterfactual case of complete isolation. Thus another way of expressing the metaphysical picture associated with the open systems view is to see it as standing in opposition to whole-to-part reductionism: ontically, it involves the assertion that the universe is not made up of small autonomous parts which behave like autonomous closed systems in between interactions, and methodologically it suggests that the appropriate way to describe reality is not to build it up out of individual closed systems, but rather to learn how to work with open systems.

In this chapter we will employ the following definition of 'closure', taken from Cuffaro and Hartmann (2024): 'when the influence of the rest of the cosmos, i.e., of the environment, on a particular system of interest, *S*, is negligible in the context of a particular investigation, then it is legitimate, for the purposes of that investigation, to treat *S* as closed' (Cuffaro and Hartmann 2024: 2). Thus we will take it that a system is spatially closed only in the limit as the influence of the environment on the system goes to zero. Note that the term 'influence' is quite broad, but we take it that 'influence' is intended to connote a relation that has some kind of modal or nomic force, i.e., it results in correlations which are not merely accidental. Evidently causation is one possible kind of 'influence', but particularly in discussing temporal closure we should not assume that causation is the only possible kind of influence, since causation is usually assumed to be asymmetric and to be oriented forwards in time, whereas in the temporal case we may wish to countenance more general possibilities—for example, Adlam (2023) discusses a symmetric generalisation of causation, and Adlam (2022b) discusses time-symmetric influences in an

all-at-once picture. In this chapter we will follow Cuffaro and Hartmann in using the term 'influence' in order to accommodate this broad range of possibilities.

2.1.1 The Temporal Analogue

The practice of reducing wholes to parts also has a temporal analogue—an ontological reductionism which regards the world as being made up of separable 'time-slices', and correspondingly a methodological reductionism which enjoins us to do science by separating histories into their constituent temporal parts, studying the autonomous behaviour of the temporal parts, and thereby deducing the whole history by passing sequentially through temporal parts. This drives an evolution-based way of thinking about physics, where we evolve a time-slice to the next time-slice and then add another time-slice, and so on, making our history longer and longer until we eventually arrive at the whole of history. We will henceforth refer to both this ontological position and its methodological analogue as 'history-to-slice reductionism'.

Just as the open systems view can be regarded as standing in opposition to whole-to-part reductionism, we can envision a temporal version of the open systems view which stands in opposition to history-to-slice reductionism. Following the definition in the previous section let us say that a time-slice is temporally closed in the limit as the influence of events at other times goes to zero. Then the 'temporally open systems view' (TOSV) would suggest that systems are generically temporally *open*. Ontically, it would involve the assertion that the universe is not made up of autonomous time-slices whose instantaneous evolution is always independent from what happens in other time-slices, and methodologically it would suggest that the appropriate way to describe reality is not to build it up out of individual time-slices, but rather to learn how to work with temporally open systems.

Now, one might worry that the TOSV is trivial. After all, almost no one would deny that the past has an influence on the future, and there are even some who would argue that the future influences the past, so what is so novel about the TOSV? Well, in the temporally closed picture, events at one time influence events at another time only via the mediation of an intervening state, and thus the happenings on a given time-slice depend entirely on the state at that time-slice (which of course may contain information about the past, and/or the future). Therefore time-slices are autonomous in the sense that, given a time-slice, the laws specifying its evolution take as their argument only the state on the slice—they don't require any *additional* information about happenings in either the future or the past. Of course, in an indeterministic theory, it will usually be the case that having additional information about the future would allow us to make better predictions than we could make based only on the state on the slice, but according to the usual

way of thinking this is not because the evolution actually *depends* on events in the future, but merely because looking into the future allows us to 'cheat' by finding out the outcomes of events which are genuinely indeterministic from the point of view of the underlying dynamical processes.[1]

The TOSV, on the other hand, insists that time-slices are not autonomous and therefore happenings on a time-slice may be influenced by past or future events which are not recorded in the physical state on that time-slice, so it presents us with quite a different picture of the relations between different time-slices. Of course, since the notion of 'influence' used here is quite broad, there are many different ways in which the TOSV could potentially be implemented. Thus this proposal should be understood as having a somewhat programmatic character, with the aim of motivating a range of specific foundational approaches which may be regarded as implementing the TOSV in various different ways.[2]

Of course, at a mathematical level it's hard to give a conclusive criterion which distinguishes between models in which current events really *depend* on events in the future, and models in which current events don't depend on the future, but nonetheless looking into the future allows us to find out the outcomes of indeterministic events. And similarly it's hard to give a conclusive criterion which distinguishes between models in which current events depend directly on events in the past which are not recorded in the current state, and models in which we simply stipulate that the current state includes records of everything that has happened in the past, some of which may be stored in ways that make them inaccessible to observers like us. So to some degree the distinction between temporally closed and temporally open is interpretative rather than formal. However, the really substantive difference between the views appears when we consider their methodological implications, for adherents of the temporally closed approach will naturally feel that the overarching goal of physics is to arrive at Markovian time-evolution laws codifying the 'real' dynamics which take as their argument only states on time-slices, whereas the TOSV tells us there is no reason to think the world is fundamentally described by Markovian time-evolution laws, and thus we shouldn't limit our search for possible theories to those which naturally take a Markovian time-evolution form. Here, we are using the term 'Markovian' to refer to laws having the property that the evolution at a given time depends only on the state at that time, and not on any facts about past states—this is a natural generalisation of the term as it is commonly used in the study of stochastic processes (Stroock 2013). Thus although it may be the case that individual models can't always be straightforwardly classified as temporally open or temporally closed, nonetheless the two views suggest quite different approaches to scientific practice.

As discussed by Cuffaro and Hartmann (2024), there exists a fairly well-established mathematical framework making use of the Lindblad equation in which we can describe open systems in the quantum-mechanical context. There does not as yet exist any such general framework for the description of *temporally*

open quantum systems. However, there certainly exist specific interpretations of quantum mechanics which can be interpreted as postulating temporally open systems—for example, the relativistic GRW flash model (Tumulka 2006, Tumulka 2021), Barandes' stochastic–quantum correspondence (Barandes 2023), or Kent's solution to the Lorentzian quantum reality problem (Kent 2014, Kent 2015, Kent 2017). Recent work on indefinite causal order (Oreshkov et al. 2012, Rubino et al. 2016, Oreshkov and Giarmatzi 2016) also appears to be moving in this direction. Perhaps these kinds of models might serve as a starting point for the creation of a general framework for temporally open systems, but for now we will seek to understand the notion at a conceptual level without appealing to a specific mathematical formalism.

2.1.2 Relativistic Considerations

Evidently spatial and temporal closure are closely related, and in fact, in the context of relativistic physics it seems quite hard to insist on the temporally closed view while supporting the (spatially) open systems view. For as noted by Cuffaro and Hartmann (2024), the phenomenon of entanglement is one of the main reasons to call into question the idea that systems can be fully isolated from their surroundings. This is because one common way of thinking about entanglement suggests that given two systems which are entangled, a measurement on one system exerts some kind of influence on the other system, even though nothing passes between them to mediate that influence. This means that shielding a system from physical exchanges of matter and energy cannot in general prevent it from being influenced by its surroundings, and therefore we should regard physical systems as being generically open rather than closed. Then note that if we think of this unmediated influence as taking place instantaneously, in a relativistic setting we can always choose some other reference frame in which this unmediated influence takes place not instantaneously but instead across some temporal interval. So, in a relativistic setting, as soon as we allow the possibility of unmediated influences across spatial distances, we must also allow that in some reference frames we have unmediated influences across temporal intervals, and therefore generic spatial openness appears to lead to generic temporal openness as well.

Thus in addition to defining spatial and temporal closure separately, it is helpful to also define their combination—a system can be regarded as *spatiotemporally* closed as the influence on it of systems and events spatially and/or temporally separated from it goes to zero. The combination of spatial and temporal closure is thus closely related to 'spatiotemporal separability', i.e., the principle that the properties of a spacetime region supervene entirely on the properties of its spatiotemporal parts (Healey and Gomes 2022). It is also close to Butterfield's pointillisme as regards spacetime. So the above considerations suggest that as soon

as we deny either spatial or temporal closure individually, we are liable to be led towards denying spatiotemporal closure as a whole, and hence towards denying spatiotemporal separability or pointillisme.

Now, this kind of argument is suggestive, but should be made more formal. In the following sections we aim to achieve this by applying the notion of fundamentality introduced in Cuffaro and Hartmann (2024), arguing that temporally open systems are more ontically fundamental than temporally closed ones. We thus conclude, in close analogy with the argumentative strategy of Cuffaro and Hartmann (2024), that the *temporally* open systems view should be considered more fundamental than the temporally closed systems view in quantum mechanics.

2.2 Ontic Fundamentality

Cuffaro and Hartmann define ontic fundamentality as follows: 'if two frameworks F_1, F_2 with fundamental Objects $\{O_1\}$, $\{O_2\}$ are motivated by the metaphysical positions M_1, M_2 respectively, then F_1 is ontically more fundamental than F_2 iff the way that the $\{O_2\}$ actually represent their "little-o" objects corresponds not to M_2 but to M_1.' Here, '"little-o" objects' refers to the pre-theoretic sense of the word 'object', while 'Objects' refers to those features of the theoretical framework such that the '"little-o" objects' are representable in terms of O in every model of every theory that can be formulated in F.

In this section, we will argue that just as the open systems view turns out to be ontically more fundamental than the closed systems view, so the TOSV turns out to be ontically more fundamental than a view on which systems are typically temporally closed. Our first metaphysical position M_1 is thus a view on which systems are typically temporally closed, i.e., history-to-slice reductionism, and our second metaphysical position M_2 is the TOSV. As for the theoretical frameworks, the framework F_1 associated with the temporally closed view can be taken as either standard quantum theory (ST) or the general quantum theory of open systems (GT), since at least superficially both of these frameworks describe systems in a way that seems to adhere to the temporally closed systems view. Even though GT explicitly postulates that the universe does not break down into spatially small pieces which behave autonomously, it maintains at least the veneer of a metaphysical picture in which we can split the universe into temporally small parts (i.e., time-slices) which behave autonomously. This is clear from the fact that both ST and GT postulate Markovian dynamical equations—the Schrödinger equation in the case of ST, and the Lindblad equation in the case of GT (indeed, as explained by Cuffaro and Hartmann (2024), Markovianity is one of the key assumptions used when we derive the Lindblad equation within the open systems picture).

Meanwhile, as noted above, there does not currently exist any general theoretical framework designed for the TOSV picture. There are some options

we might consider, including path integral formulations of quantum mechanics (Feynman et al. 2010, Wharton 2016), operational theory approaches (Abramsky and Heunen 2013, Oreshkov and Cerf 2016), and the process matrix framework (Oreshkov and Giarmatzi 2016, Oreshkov et al. 2012), but in all these cases there is some uncertainty around the ontological commitments of the framework, and thus it would be premature to declare any of them as the correct framework F_2 to represent the TOSV. However, to apply the definition of ontological fundamentality suggested by Cuffaro and Hartmann we don't actually have to know anything at all about the framework F_2; we need only examine the framework F_1 and compare it to the two motivating metaphysical positions. So we can conclude that framework F_2 is more ontically fundamental than framework F_1 even without actually coming up with any concrete formulation of framework F_2, since ultimately what counts for this definition is the motivating metaphysical position. Indeed, it is to be hoped that by explicitly recognising the fundamentality of the TOSV we might stimulate further work on the development of an appropriate theoretical framework for it.

Let us therefore take a closer look at how the objects of ST and GT represent their little-o objects. First, the Objects are state vectors in ST and density operators in GT, which each represent the little-o object known as 'the state of the system'. Since both state vectors and density operators are defined on a single time-slice, at first glance it does indeed seem that both ST and GT represent their little-o objects in a way that corresponds to the temporally closed view.

However, because the temporally closed approach insists that time-slices are fundamentally distinct and autonomous, it follows from this view that specifying the state on every time-slice in some chosen foliation should be sufficient to fully specify all the properties of the entire history of the system. In particular, this suggests that specifying the state on every time-slice in some foliation should be sufficient to specify the state on every time-slice in any other foliation—a theory with this property is known as narratable. Note that the argument for narratability does not rely on the relativistic principle that there exists no privileged foliation, for even if there *does* exist some privileged foliation of a history which encodes an ontologically privileged decomposition of that history into time-slices, in a temporally closed view it ought to be the case that specifying all of the information about all of the states on the time-slices in that foliation exhausts all of the information about the properties of the history, so we ought to be able to reconstruct any other possible foliation from that complete information. Thus the temporally closed view on its own already implies narratability, without any appeal to relativistic ideas.

But it is known that in general, quantum mechanics is *not* narratable: 'a state history of the composite system given along one foliation might not, by itself—that is, without specification of the laws governing the dynamical evolution of the systems—determine the state history along another' (Myrvold 2016: 761; see Albert 2015). Thus since ST and GT are both versions of quantum mechanics,

neither of them obeys narratability—and therefore, despite appearances, they must actually represent their little-o objects in a way that corresponds to the TOSV, which indicates that the TOSV must be regarded as more ontically fundamental than the temporally closed approach.

The origin of the failure of narratability in quantum mechanics is straightforward: in quantum mechanics (and thus in both ST and GT), the state of a composite system composed of several entangled parts is underdetermined by the states of its composite parts. Now suppose two systems A and B are entangled, with a state ρ_{AB} on some hypersurface H and another state ρ'_{AB} on some other hypersurface H'. If we change our foliation such that we are now dealing with a hypersurface G such that A coincides with the intersection of G and H, and B coincides with the intersection of G and H', we can then infer that the reduced states of A and B on G are respectively $Tr_B(\rho_{AB})$ and $Tr_A(\rho'_{AB})$, but this is not enough to infer the joint state of A, B on the hypersurface G, so unless we have some additional information about the dynamics between H and H' we won't be able to infer the full state on G. Note that the failure of narratability arises specifically when we are dealing with states that are not separable—which of course are made up of two or more open systems (open in the sense that each system is influenced by the other systems with which it is entangled). When we write down the history of a set of open systems, in general we will find that 'the piecewise construction of a state on one hypersurface out of bits of states on hypersurfaces that intersect it will not work' (Myrvold 2016). That is to say, once we accept that we are typically dealing with open quantum systems rather than closed ones, it follows that history-to-slice reductionism will generically fail. A universe made up principally of open systems cannot in general be understood as a set of time-slices which together exhaust the content of reality: just as that universe is not composed of autonomous spatial parts, so it is also not composed of autonomous temporal parts.

We note that, as shown by Myrvold (2016), the failure of narratability applies specifically to bipartite systems in pure entangled states—for larger systems, imposing relativistic causality constraints is usually enough to ensure that specifying the state at all times is sufficient to specify a state history on any other foliation. But as emphasised by Myrvold, this is still a significant difference from the classical case, where we expect to be able to go from one foliation to another without placing any constraints at all on the dynamics. So this point does not undermine our contention that ST and GT cannot in general represent their objects in such a way that specifying the state at all times is sufficient to specify the entire history of the system. The fact that further dynamical constraints must be applied to infer the full description from the time-slice description means precisely that the time-slices cannot be regarded as fundamentally distinct and autonomous: they are subject to global constraints which undermine their supposed autonomy and therefore the decomposition into temporal time-slices does not give the whole story in the way that it does in a classical context.

2.2.1 Interpretation

The failure of narratability is a clear sign that decomposing quantum mechanics into timelike slices is the wrong way of thinking about its ontology. But to say more about ontic fundamentality in quantum mechanics, it is helpful to look at some specific interpretations of the theory, so we will now consider how history-to-slice reductionism fares in the context of the Everett interpretation and the orthodox interpretations (we follow Cuffaro and Hartmann in focusing on unitary-only approaches to quantum mechanics, since the non-unitary approaches are really distinct theories and thus will not necessarily give the same conclusions about fundamentality).

The Everett Interpretation

The Everett interpretation (Everett 1957, Wallace 2012) tells us that unitary quantum mechanics is universal and complete. In the decoherence-first Everett approach it is then argued that decoherence leads to the emergence of high-level branching structure—so for example, when a macroscopic event that we would characterise as a quantum measurement is performed, all possible measurement outcomes always occur, but different outcomes are witnessed by distinct versions of the relevant observer in separate emergent branches of the wavefunction, so that each version of the observer is only aware of a single outcome.

As noted by Cuffaro and Hartmann, the Everett interpretation clearly posits some open systems, although it also posits closed systems. But does the same hold for *temporal* closure? The immediate answer may seem to be negative: the ontology of the Everett interpretation includes nothing other than the quantum state undergoing unitary evolution, and since unitary evolution can always be decomposed onto autonomous spacelike hypersurfaces, it would seem that all the systems in the Everett interpretation must be temporally closed. But in fact, the ontology of the Everett interpretation is not as straightforward as it would seem—for recall that the wavefunction does not live on ordinary spacetime, so the Everett interpretation does not, in fact, present us with a picture in which the quantum state simply evolves *on* spacetime. In fact 'the space on which the wavefunction lives is not familiar 3-dimensional space (represented mathematically by $\mathbb{R}^3$) but rather 3N-dimensional space ("configuration space" as it is called in physics)' (Wallace 2017). What ontology are we to associate with this?

Some Everettians have suggested an ontology known as *wavefunction realism*, where the state of the whole system is given by the separate wavefunction values at each point of configuration space (Albert 1996, Monton 2006, Lewis 2004). The state thus evolves in a space which is composed of configuration space plus a temporal dimension. This might seem to support the methodology of history-to-slice reductionism, as the history is indeed divided into autonomous slices, each represented by a wavefunction over configuration space. However, the evolution

relating these points takes place in configuration space rather than real space, and it isn't clear that the division into autonomous slices of configuration space can be straightforwardly mapped to a set of autonomous slices of ordinary spacetime. Moreover, if we take into account the relativistic idea that there are no preferred reference frames, it follows that the division into configuration space plus a temporal dimension cannot be taken entirely literally: if we choose a different foliation, we will end up with a different path through configuration space which nonetheless describes the same spatiotemporal history. Thus a relativistic version of this approach would presumably have to work with something like an equivalence class of paths through configuration space, and such an ontology does not seem to correspond literally to autonomous time-slices even in configuration space. So there is a case to be made that in wavefunction realism, the 'little-o' objects are really represented in a way corresponding to the TOSV rather than the temporally closed view, making the TOSV more ontically fundamental.

Moreover, wavefunction realism has a number of serious conceptual and technical problems (Wallace 2017): as explained by Wallace and Timpson (2009), 'it is plausible only in the restricted setting of fixed-particle-number non-relativistic quantum mechanics and it has no reasonable extension to more physically realistic situations'. Wallace and Timpson have therefore proposed an alternative ontology known as *spacetime state realism*, which associates a set of properties (represented by a density operator) to each region of spacetime. Note that Wallace and Timpson start from an approach which associates states with regions of *space*, but as they acknowledge, once we move to full quantum field theory it becomes inevitable that the states should instead be associated with small regions of *spacetime*—for example, this follows from 'the algebraic formulation of QFT, (in which) we associate to each spacetime region R a C^*-algebra $A(R)$ of operators, representing the dynamical variables associated to region R' (Wallace and Timpson 2009: 711). Indeed, a number of features of QFT serve to problematise the idea of instantaneous states on spatial slices: as Papageorgiou and Fraser (2023) put it, 'Algebraic states are time-independent; they are associated with spacetime regions indirectly, as functionals of algebras directly associated with spacetime regions … Physically significant states … are global states in the sense of being states for all of space and time … for any algebraic state ω a time evolution cannot be introduced by restrictions of ω to a foliation of Cauchy surfaces.' So the move made by Wallace and Timpson seems like a natural one in the context of QFT.

But it must be reinforced that this move has significant consequences for the ontology of the resulting theory. For if we accept this view, it follows that the fundamental objects of the Everett interpretation typically have finite *temporal* extent, and thus even if these regions were generically 'closed', nonetheless their non-zero temporal extent would make strict history-to-slice reductionism impossible. Thus there are immediately reasons to think that in this picture the 'little-o' objects are ultimately being represented in a way corresponding to the TOSV rather than a

temporally closed approach. Moreover, Wallace and Timpson note that in general the states of these regions will be density matrices rather than pure quantum states, which is to say that they are *not* generically closed, i.e., they are subject to external influences. And since these regions typically have nonzero temporal extent in any given reference frame, it follows that they are not only spatially but also *temporally* open, since no event will be simultaneous with the whole region in any given reference frame. So again, the representation seems to correspond to the TOSV rather than a temporally closed approach, making the TOSV more ontically fundamental.

Now, the proponent of the closed systems view might try to argue that the spacetime regions proposed by Wallace and Timpson can be fundamental only if their state is pure; so if we find a region whose state is not pure, we should expand the size of the region until we find a pure state, which we can then regard as a closed system. However, there is no obvious reason to think that we will find an absolutely pure state for any region smaller than the whole of space and time. And Wallace and Timpson emphasise that it is important that the theory focuses on subregions rather than on the whole of the cosmos: 'if, by contrast, we were to treat the Universe just as one big system, with no subsystem decomposition, then we would only have a single property bearer (the Universe as a whole) instantiating a single property (represented by the Universal density operator) and we would lack sufficient articulation to make clear physical meaning of what was presented' (Wallace and Timpson 2009: 710). Since subregions will not as a general rule be in pure states, this argument implies that if we are to have enough structure to represent the universe in a meaningful way, we must adopt a description whose fundamental objects are spatially and temporally open systems, and thus the TOSV must be regarded as more ontically fundamental.

Now, we acknowledge that the considerations which lead Wallace and Timpson to prefer spacetime state realism and to define the states on spacetime regions rather than spacetime slices are derived from relativistic quantum mechanics and quantum field theory, so one might perhaps argue that we can still employ wavefunction realism or a form of 'spatial state realism' if our aim is to understand the ontology implied by standard non-relativistic quantum mechanics. But if we look purely at non-relativistic quantum mechanics, it is not obvious how we can choose between wavefunction realism and spacetime state realism, so it seems that no particular ontology is unambiguously implied by the non-relativistic theory. And in any case, it doesn't make much sense to ignore relativistic considerations whilst trying to interpret quantum mechanics; if we do, we will only have to throw all our conclusions away when we move to relativistic quantum mechanics and QFT. The interpretation of quantum mechanics must be informed by subsequent developments in physics, and therefore it is appropriate for us to keep in mind the considerations raised by Wallace and Timpson in arguing for the fundamentality of the TOSV.[3]

There is a more general message to be taken from this saga. For the problems associated with wavefunction realism arise to some extent from its attempt to decompose evolution into autonomous time-slices, because this approach is in tension with the relativistic setting of QFT. Spacetime state realism solves these problems precisely by ceasing the attempt to decompose the ontology into temporally closed parts. And similar stories appear in the context of many other approaches to quantum mechanics: for example, Esfeld and Gisin (2013) have noted that the relativistic GRW approach to the measurement problem is relativistically covariant only if 'one renounces an account of the temporal development of the actual distribution of the flashes in space-time'—that is to say, we have to think of the distribution of flashes as being generated 'all-at-once' in a temporally open way, rather than decomposed into autonomous time-slices. We conjecture that this is a universal feature: in general, ontologies for quantum mechanics made up of autonomous temporal parts are very difficult to reconcile with relativity, and renouncing temporal closure is often a route to solving these problems.

Orthodox Interpretations

Orthodox interpretations of quantum mechanics are characterised by the following claims: a) unitary quantum mechanics is universal and complete, b) relative to each observer, every measurement has a unique outcome, and c) quantum mechanics does not describe systems as having observer-independent states. Some views which may possibly be interpreted as falling into this category include Copenhagen interpretations (Bohr 1987, Heisenberg 1958, Pauli et al. 1994), neo-Copenhagen interpretations (Zeilinger 1999, 2002, Brukner 2017, Bub 2012, Demopoulos 2012, Janas et al. 2021, de Muynck 2004), QBism (Fuchs et al. 2014), pragmatist interpretations (Healey 2012) and some versions of relational quantum mechanics (Rovelli 1996)

Rather than regarding quantum states as intrinsic properties of the system to which they are ascribed, orthodox interpretations tell us that 'what the state vector compactly encodes ... is a collection of unitarily related conditional probability distributions through which one can effectively characterise a given system in the context of a particular physical interaction' (Cuffaro and Hartmann, this volume). Thus according to the orthodox interpretations, quantum mechanics describes only what will occur in physical interactions; it does not tell us about anything that is happening in between physical interactions. Moreover, if the quantum state is only well-defined relative to an observer, its time evolution must also be relativised to an observer; this was shown formally in Fuchs (2002), and similarly Barzegar (2020) notes that 'According to QBism, just as quantum states are personal judgments, the unitary time evolutions are personal judgments too.' Therefore in the orthodox interpretations, time evolution must be understood as a process by which observers update the probability distributions characterising the system, rather than as a real physical process taking place in spacetime—or at least, if it is a real

physical process taking place in spacetime, there are many different versions of it taking place relative to different observers.

This suggests that in an orthodox interpretation, a decomposition of the history of a quantum system into time-slices related by unitary evolution cannot be taken too seriously; this decomposition is simply one possible representation of the update rule for the probability distributions describing physical interactions, and these probability distributions are the real content of quantum mechanics. So although these interpretations may still use the unitary formalism when convenient, nonetheless at an ontological level they tell us that when an interaction takes place on some time-slice, it is not the case that the outcome of the interaction is produced by dynamical evolution from the state on the immediately preceding time-slice; really that interaction is related *directly* to past interactions, without mediation by intervening states on intervening time-slices. It follows that although ST and GT might appear to present us with a decomposition into a set of autonomous time-slices, the orthodox interpretations tell us not to take this decomposition literally, and thus the way they represent their little-o objects really corresponds to the TOSV rather than a temporally closed view. So it seems reasonable to say that in an orthodox interpretation, the TOSV is more ontically fundamental than the temporally closed view, and indeed it is quite difficult to make sense of the temporally closed view at all in such a setting.

2.3 The State of the Universe

The possible failure of history-to-slice reductionism is relevant to the question raised by Cuffaro and Hartmann about whether we should expect the universe as a whole to behave as a closed system (i.e., its state is pure and it evolves unitarily) or an open system (i.e., its state may be mixed and it may evolve non-unitarily). For as Cuffaro and Hartmann (2024) argue, since gravity and entanglement cannot be shielded, 'there are reasons stemming from physics to believe that the cosmos is the only closed system that truly exists' (Cuffaro and Hartmann 2024: 2). So in the face of generic openness for subsystems of the universe, the last refuge of the closed systems view is to insist that the cosmos is a closed system and the behaviour of all the individual parts are derived from the unitary evolution of the pure state of the cosmos.

But what is meant by 'the universe as a whole' here? If this term is supposed to refer to 'the whole of the universe *at a given time*' (i.e., on some spacelike hyperplane) then the 'privileged ontological status' is not so clear. For while the whole of the universe on a given spacelike hyperplane is of course spatially closed, there is still something outside of it *temporally*, and it is certainly possible at least in principle that events on the hyperplane could be influenced directly by events on other hyperplanes, so there is no particular reason to think that the whole of the

universe on some hyperplane is subject to no external influences whatsoever. Thus if we allow influences across times, it is not inevitable that the universe at a given time must be a closed system. In this picture, there is simply no type of system which can be guaranteed to behave autonomously—the universe is not made up of autonomous parts, either spatially or temporally.

Of course, this is not to say that there could not be instances where some part of the universe behaves in a way that is approximately autonomous for some period of time: various kinds of correlations with other parts of the universe may cancel out, or may simply be irrelevant in some particular context. But there is no reason to expect autonomy in general; the fundamental stuff of reality is not a set of autonomous systems *or* a set of autonomous time-slices.

Now, the proponent of a closed system approach might perhaps try to insist that what is 'closed' is really the *whole* cosmos, i.e., not just a single time-slice but the whole of space and time. And this seems correct, but also vacuous, because 'the whole of space and time' has no dynamics: change happens *in* time, and there is no time external to 'the whole of space and time' in which the *whole* cosmos could undergo change. Thus we clearly cannot contend that the behaviour of all of the individual parts of the cosmos are derived from the autonomous dynamics of the whole cosmos, because it doesn't have any dynamics at all. Thus retreating to the notion that the whole history of the universe is the only closed system is tantamount to giving up on any meaningful closed systems view.

The proponent of the closed systems view might also try to argue as follows: given a set of quantum systems, provided that the spacetime on which they live is globally hyperbolic, we can always choose a preferred foliation of spacetime on which to describe their evolution. And this description will represent happenings such that acting on one part of an entangled system (e.g., by measurement) will always cause a change in the other part of the entangled system which is instantaneous relative to the preferred foliation. So in this description, no system is ever influenced by events taking place at times other than the present time (as defined by the chosen foliation), and thus a description of this kind does indeed represent the universe as a whole as temporally closed. Thus the proponent of the closed systems view might maintain that what they mean by saying that the universe as a whole is 'closed' is either that it is possible to choose *some* foliation of spacetime on which all influences can be treated as instantaneous, or that for *any* possible foliation of spacetime all influences can be treated as instantaneous.

However, there are a number of problems with this account. For a start, we don't know that our spacetime is in fact globally hyperbolic, and if it is not then no such foliation will be possible. Furthermore, it is important to keep in mind that modes of description which have succeeded for subsystems of the universe will not necessarily succeed for the universe as a whole. In particular, when we consider a scenario where measurements are performed on a spacelike separated entangled pair, part of the reason why we are always able to find a description in which all

influences are instantaneous is precisely the fact that we are dealing with two distinct subsystems which exist continuously, and therefore if the measurement on particle *A* has some kind of influence on a later measurement on particle *B*, we can always choose a description such that the measurement on particle *A* exerts its influence on particle *B* instantaneously, since particle *B* exists at the time of the measurement on particle *A* in any reference frame. But if the system we are dealing with is 'the entire universe', and we are interested in the question of whether events on one time-slice may have a direct influence on events on other time-slices, evidently we do not have the option of simply redescribing the situation such that all influences are instantaneous, since two distinct copies of 'the whole universe' never exist at the same time. Thus it cannot necessarily be taken for granted that in this case we will still be able to find a description in which all influences are instantaneous.

Moreover, many physicists believe on the basis of relativistic physics that the universe should not have any preferred reference frames, and if that is correct, then even if it is possible to find some way of describing the evolution of the universe such that it appears temporally closed relative to some particular foliation, this description cannot be taken too literally: it's not clear that anything at all can meaningfully be said about 'when' the influences in question take place, but in any case if there are no preferred reference frames these influences certainly cannot be literally instantaneous according to some particular foliation, since that would constitute a preferred reference frame. Thus coming up with a foliation on which all influences are instantaneous may be mathematically convenient, but it does not follow that we should accept a metaphysical picture which tells us that the slices in question are actually temporally closed.

2.4 Conclusion

We have argued that a move towards the open systems view as advocated by Cuffaro and Hartmann should be accompanied by a move towards a *temporally* open systems view: for just as there are good reasons to think of the open systems view as more ontically fundamental than the closed systems view, a similar analysis suggests that the TOSV is more ontically fundamental than a temporally closed view.

We have also argued that the fundamentality of the TOSV has important implications for the questions raised by Cuffaro and Hartmann about whether the state of the universe as a whole is open or closed. For if temporal closure is not generic, this means we cannot take for granted that even the whole universe at a single time is really 'closed'—only the whole of spacetime can be regarded as 'closed', and that statement is trivial, since the whole of spacetime does not undergo change or evolution. This provides further justification for the hypothesis that we should not

expect the universe as a whole to behave as a closed system, and thus the possibility of the TOSV should be taken into account in this debate.

Finally, our application of the definition of ontic fundamentality suggested by Cuffaro and Hartmann to the question of temporal closure demonstrates the general applicability of this definition outside of the original context in which it was proposed, suggesting that this definition may also be valuable in other long-standing discussions about fundamentality.

Endnotes

1. Note that the notion of 'determinism' becomes complex in the context of the TOSV or any view which repudiates fundamental time evolution—see Adlam (2022a) for further details.
2. Thanks to an anonymous reviewer for raising this point.
3. With this in mind, we note that it seems unlikely that spacetime state realism will survive intact in a theory of quantum gravity—but that lies beyond the scope of this chapter.

Bibliography

Abramsky, S., and Heunen, C. (2013), 'Operational Theories and Categorical Quantum Mechanics'. arXiv:1206.0921v4.

Adlam, E. (2022a), 'Determinism Beyond Time Evolution', in *European Journal for Philosophy of Science*, 12: 73.

Adlam, E. (2022b), 'Two Roads to Retrocausality', in *Synthese* 200: 422.

Adlam, E. (2023), 'Is There Causation in Fundamental Physics? New Insights from Process Matrices and Quantum Causal Modelling', in *Synthese* 201: 5:1–40.

Albert, D. Z. (2015), 'Physics and Narrative', in *After Physics* (Cambridge, MA: Harvard University Press), 106–23.

Albert, D. Z. (1996), 'Elementary Quantum Metaphysics', in T. J. Cushing and A. Fine (eds.), *Bohmian Mechanics and Quantum Theory: An Appraisal*. Boston Studies in the Philosophy of Science, vol. 184 (Dordrecht: Springer), 277–84.

Barandes, J. A. (2023), 'The Stochastic–Quantum Correspondence'. *Philosophy of Physics* 3(1): 8.

Barzegar, A. (2020), 'QBism Is Not So Simply Dismissed', in *Foundations of Physics* 50: 693–707.

Bohr, N. (1987), *The Philosophical Writings of Niels Bohr* (Woodbridge, CT: Ox Bow Press).

Brukner, Č. (2017), 'On the Quantum Measurement Problem', in *Quantum [Un]Speakables* II (Springer), 95–117.

Bub, J. (2012), *Bananaworld: Quantum Mechanics for Primates* (Oxford: Oxford University Press).

Butterfield, J. (2006), 'Against Pointillisme about Mechanics', in *The British Journal for the Philosophy of Science* 57: 709–53.

Castellani, E. (2002), 'Reductionism, Emergence, and Effective Field Theories', in *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 33: 251–67.

Cuffaro, M. E., and Hartmann, S. (2024), 'The Open Systems View', in *Philosophy of Physics* 2: 6:1–27.

de Muynck, W. M. (2004), 'Towards a Neo-Copenhagen Interpretation of Quantum Mechanics', *Foundations of Physics* 34: 717–70.
Demopoulos, W. (2012), 'Generalized Probability Measures and the Framework of Effects', in Y. Ben-Menahem and M. Hemmo (eds.), *Probability in Physics* (Heidelberg: Springer), 201–17.
Esfeld, M., and Gisin, N. (2013), 'The GRW Flash Theory: A Relativistic Quantum Ontology of Matter in Space-time?'. arXiv:1310.5308v1.
Everett, H., III (1957), '"Relative State" Formulation of Quantum Mechanics', in *Reviews of Modern Physics* 29: 454–62.
Feynman, R. P., Hibbs, A. R., and D. F. Styer (2010), *Quantum Mechanics and Path Integrals* (Mineola: Dover Publications).
Fuchs, C. A. (2002), 'Quantum Mechanics as Quantum Information (and Only a Little More)'. arXiv:quant-ph/0205039v1.
Fuchs, C. A., Mermin, N. D., and R. Schack (2014), 'An Introduction to QBism with an Application to the Locality of Quantum Mechanics', in *American Journal of Physics* 82: 749–54.
Healey, R. (2012), 'Quantum Theory: A Pragmatist Approach', in *British Journal for the Philosophy of Science* 63: 729–71.
Healey, R. and Gomes, H. (2022), 'Holism and Nonseparability in Physics', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University), Spring 2022 edition. https://plato.stanford.edu/archives/spr2022/entries/physics-holism/
Heisenberg, W. (1958), *Physics and Philosophy: The Revolution in Modern Science* (New York: Harper).
Janas, M., Cuffaro, M. E., and Janssen, M. (2022), *Understanding Quantum Raffles: Quantum Mechanics on an Informational Approach—Structure and Interpretation* (Foreword by J. Bub) (Cham: Springer).
Kent, A. (2014), 'Solution to the Lorentzian Quantum Reality Problem', in *Physical Review A* 90: 012107.
Kent, A. (2015), 'Lorentzian Quantum Reality: Postulates and Toy Models', in *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 373: 20140241.
Kent, A. (2017), 'Quantum Reality via Late-Time Photodetection', in *Physical Review A* 96: 062121.
Lewis, P. J. (2004), 'Life in Configuration Space', in *The British Journal for the Philosophy of Science* 55: 713–29.
Monton, B. (2006), 'Quantum Mechanics and 3*n*-Dimensional Space', in *Philosophy of Science* 73: 778–89.
Myrvold, W. C. (2016), 'Quantum Mechanics and Narratability', in *Foundations of Physics* 46: 759–75.
Oreshkov, O., and Cerf, N. J. (2016), Operational Quantum Theory Without Predefined Time', in *New Journal of Physics* 18: 073037.
Oreshkov, O., Costa, F., and Brukner, Č. (2012), 'Quantum Correlations with No Causal Order', in *Nature Communications* 3: 1092.
Oreshkov, O., and Giarmatzi, C. (2016), 'Causal and Causally Separable Processes', in *New Journal of Physics* 18: 093020.
Papageorgiou, M., and Fraser, D. (2023), 'Eliminating the "Impossible": Recent Progress on Local Measurement Theory for Quantum Field Theory', arXiv:2307.08524.
Pauli, W., Enz, C. P., and von Meyenn, K. (1994), *Writings on Physics and Philosophy* (Berlin: Springer).

Rovelli, C. (1996), 'Relational Quantum Mechanics', in *International Journal of Theoretical Physics* 35:1637–78.

Rubino, G., Rozema, L., Feix, A., Araujo, M., Zeuner, J., Procopio, L., Brukner, Č., and Walther, P. (2016), 'Experimental Verification of an Indefinite Causal Order', in *Science Advances* 3: e1602589.

Stroock, D. W. (2013), *An Introduction to Markov Processes*, Graduate Texts in Mathematics (Berlin: Springer).

Tumulka, R. (2006), 'A Relativistic Version of the Ghirardi–Rimini–Weber Model', in *Journal of Statistical Physics* 125: 821–40.

Tumulka, R. (2021), 'A Relativistic GRW Flash Process with Interaction', in V. Allori, A. Bassi, D. Dürr, and N. Zanghi (eds.), *Do Wave Functions Jump?* Fundamental Theories of Physics, vol. 198 (Cham: Springer), 321–47.

Wallace, D. (2017), 'Against Wavefunction Realism', in S. Dasgupta, R. Dotan, and B. Weslake (eds.), Current Controversies in Philosophy of Science (New York: Routledge), 63–74.

Wallace, D., and Timpson, C. G. (2010), 'Quantum Mechanics on Spacetime I: Spacetime State Realism', in *The British Journal for the Philosophy of Science* 61: 697–727.

Wallace, W. (2012), *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation* (Oxford: Oxford University Press).

Weinberg, S. (2001), '10. Reductionism Redux', in *Facing Up* (Cambridge: Harvard University Press), 107–22.

Wharton, K. (2016), 'Towards a Realistic Parsing of the Feynman Path Integral', in *Quanta* 5: 1–11.

Zeilinger, A. (1999), 'A Foundational Principle for Quantum Mechanics', in *Foundations of Physics*, 29: 631–43.

Zeilinger, A. (2002), 'Bell's Theorem, Information and Quantum Physics', in R. A. Bertlmann, and A. Zeilinger (eds.), *Quantum [Un] Speakables* (Berlin: Springer), 241–54.

3
The Relative Nature of Open Quantum Systems

Olimpia Lombardi

3.1 Introduction

Unlike biology, physics has traditionally adopted a closed systems view, and standard quantum physics (ST) is not an exception in this regard. However, ST is very peculiar. Using Einstein's words (1948, translated by Howard 1985: 187), 'the mutually independent existence (the "being-thus") of spatially distant things' is in general assumed in non-quantum physics; so, 'things' can be individualised at least by their space-time trajectories. In contrast, precise trajectories cannot be easily assigned to quantum systems, and quantum correlations challenge the independent existence of interacting systems. Therefore, in the quantum framework the identification of open systems is by no means a trivial matter.

In this chapter, I will emphasise the relative nature of open quantum systems in ST by pointing out how this is manifested under different forms. Therefore, any view that considers open quantum systems as fundamental must take this relativity into account. By contrast, the fact that open systems can only be defined relative to one another comfortably fits in a closed systems view. I will recall a top-down view of quantum mechanics introduced in previous papers, which conceives the theory as directly applied to closed systems and only derivatively applied to open systems. From this perspective, I will briefly consider some claims about the fundamentality of the open systems view (Cuffaro and Hartmann, this volume).

3.2 Open System: A Relative Notion

In the classical framework, not only can systems be identified by their space-time location, but they also have 'mutually independent existence': they are individuals in the philosophical meaning of the concept. In the quantum domain, by contrast, matters are not so easy: after interacting, systems become interrelated in a way that diminishes neither with time nor with space. Therefore, open systems cannot be

Olimpia Lombardi, *The Relative Nature of Open Quantum Systems*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press.
© Oxford University Press (2026). DOI: 10.1093/9780198929277.003.0004

univocally identified: they have an intrinsically relative nature, which manifests itself in different ways.

3.2.1 The Relativity of Entanglement

Although the term 'entanglement' is commonly applied to particles or systems, as if it denoted a relation that links quantum objects to each other, the concept requires a careful attention.

First, entanglement does not require interaction: 'entanglement swapping'—first introduced in Żukowski et al. (1993); for its first experimental realisation, see Pan et al. (1998)—shows that entanglement can be established between systems that have never physically interacted with each other. Second, there may be entanglement in a single particle. For example, intraparticle entanglement (Harshman and Wikramasekara 2007: 2) may exist between the momentum degrees of freedom and the spin degrees of freedom of a single atom (see also Harshman 2012) or between two spatial modes of a single photon passing through a beam splitter (Terra Cunha, Dunningham, and Vedral 2007). Furthermore, one-particle entanglement is as good as two-particle entanglement as a resource: a one-photon or one-atom state can also be used to violate Bell's inequalities and to teleport a qubit (Lee and Kim 2000).

Such facts show that entanglement must be conceived not as a relation between particles, but as a feature of the state of a composite system. Nevertheless, it is not an absolute feature. As John Earman (2015) clearly emphasises, *a given state is entangled or not only in relation to the decomposition of the composite system into subsystems*. In other words, a particular state may be entangled with respect to one decomposition and non-entangled with respect to another. In other words, *entanglement is an essentially relative notion*. As Terra Cunha and collaborators (2007: 2278) claim, 'entanglement is a property of a quantum state relative to a given set of subsystems … a question like "is there entanglement in such a state?" is meaningless in isolation'. In this context the authors prove a theorem according to which, given a state vector $|\psi\rangle$ in a finite-dimensional state space $\mathcal{H}$ with non-prime dimension $d = mn$, there always exists a tensor product structure $\mathcal{H} = \mathcal{H}^m \otimes \mathcal{H}^n$ with respect to which $|\psi\rangle$ is factorisable, that is, non-entangled.

In the algebraic formalism, the concept of entanglement can be generalised to the case in which the decomposition of the composite system cannot be given in terms of a tensor product structure, as in general in the bosonic and fermionic cases. For instance, Howard Barnum and collaborators (Barnum et al. 2003, 2004; Viola and Barnum 2010) introduce a generalisation based on the idea that entanglement is relative to a specified subspace of observables rather than to a specified decomposition of the composite system into subsystems. Therefore, the concept

of quantum correlation can also be generalised in such a way to be relativised to the subalgebras of the algebra of observables of the composite system (Bellomo et al. 2014).

3.2.2 The 'Defining-Systems' Problem in Decoherence

In the 'environment-induced' approach to decoherence (EID, see, e.g., Zurek 1982, 2003), the first step is to partition the closed system into some degrees of freedom that represent the system of interest, and the remaining degrees of freedom that play the role of the environment. For example, in the cosmological context, long wavelength modes are usually considered as the system, and short wavelength modes are conceived as the environment (Lombardo and Mazzitelli 1996). However, this is not the only way to introduce the split: in other cases it is supposed that the tensor and the scalar fluctuations interact with each other, and the tensor fluctuations act as an environment that causes the loss of coherence of the scalar fluctuations, whose classicality is so justified (Franco and Calzetta 2011).

One of the conceptual challenges of EID is the so-called 'defining-systems problem' (see Castagnino, Laura, and Lombardi 2007): EID does not provide a criterion to decide where to place the 'cut' between the proper system and the environment. In general, the classically behaving degrees of freedom are assumed in advance: the application of the EID formalism does not predict which observables will manifest a classical behaviour but only confirms a previous assumption. When classicality is not assumed in advance, decoherence depends on where the boundary of the open system is placed. Furthermore, in a generalisation of the spin-bath model it has been proved that, whereas certain particles decohere and may behave classically, the subsystem composed by all of them does not decohere and retains its quantum nature (see Castagnino, Fortin, and Lombardi 2010; Lombardi, Fortin, and Castagnino 2012). This problem is acknowledged by Zurek himself: 'one issue which has been often taken for granted is looming big, as a foundation of the whole decoherence program. It is the question of what are the "systems" which play such a crucial role in all the discussions of the emergent classicality. This issue was raised earlier, but the progress to date has been slow at best' (Zurek 2000: 338).

3.2.3 Quantum Structure Studies

The so-called 'quantum structural studies' deal with the different ways in which a quantum system can be decomposed into subsystems (quantum structures) (see, e.g., Harshman and Ranade 2011, Kastner, Jeknić-Dugić, and Jaroszkiewicz 2016). The typical case is that of the hydrogen atom, which can be decomposed

into the proton system and the electron system, but also into the centre-of-mass system and the relative system (see Ardenghi, Castagnino, and Lombardi 2009, Harshman 2012). Moreover, 'without further physical assumption, no partition has an ontologically superior status with respect to any other' (Zanardi 2001: 4).

From a formal viewpoint, the problem of the identification of subsystems of a composite system is expressed in terms of tensor product structures: the different ways in which a Hilbert space can be decomposed into a tensor product of Hilbert spaces. Let us consider, for example, systems with three particles with spin in one dimension (see Harshman 2016). The 'natural' tensor product structure is $\mathcal{H} = \mathcal{H}^1 \otimes \mathcal{H}^2 \otimes \mathcal{H}^3$. Already in this structure, the question arises as to which ones are the open systems: those represented in $\mathcal{H}^1 \otimes \mathcal{H}^2$ and $\mathcal{H}^3$, or those represented in $\mathcal{H}^1$ and $\mathcal{H}^2 \otimes \mathcal{H}^3$. Furthermore, the Hilbert space of each particle can be factored into a spatial part $\mathcal{K}^i$ and a spin part $\mathcal{S}^i$ as $\mathcal{H}^i = \mathcal{K}^i \otimes \mathcal{S}^i$, so that $\mathcal{H} = \mathcal{K} \otimes \mathcal{S} = \left(\otimes_i \mathcal{K}^i\right) \otimes \left(\otimes_i \mathcal{S}^i\right)$. This tensor product structure evidences the existence of open systems, represented by $\mathcal{K}$ and $\mathcal{S}$, which cannot even be identified with sets of particles.

3.3 A Top-Down Approach to Quantum Mechanics

The observations of the previous section support the idea that, in the quantum domain, open systems can only be defined relative to one another. To speak of a quantum-open system in a non-relative way makes no sense because the same degrees of freedom can constitute an open system or a different one. Therefore, the identification of subsystems of a closed system is a subtle issue: one cannot simply consider the existence of open systems without taking into account their relativity. This view is in resonance with what has been called the 'holistic' nature of quantum systems, understanding holism as the feature of some physical objects that are not composed of physical parts but are indivisible wholes (see, e.g., Healey and Gomes 2022).

These features of quantum systems have led to a 'top-down' approach to ST, according to which the theory applies primarily to closed systems and only derivatively to their constituents. Furthermore, strictly speaking only closed systems are quantum systems, because their behaviour is described by the unitary dynamical postulate of the theory. When only certain degrees of freedom are cut out of the closed system, the non-unitarily evolving result is not a subsystem because it is not a system ruled by that dynamical postulate: it is a *part* of the closed system, defined only in relation to the remaining parts. Here I will sketch the most relevant aspects of this approach.

- From this perspective, reduced states are not the quantum states of subsystems in the same sense as quantum states are states of the whole closed composite system. By relying on a generalised concept of coarse-graining, it

can be proved that reduced states are coarse-grained states of closed systems, which disregard certain degrees of freedom considered as irrelevant (Fortin and Lombardi 2014). Therefore, their non-unitary evolutions do not need the reformulation of the dynamical postulate of ST nor require a new dynamical postulate to be accounted for. The non-unitary evolution of a quantum reduced state turns out to be a situation analogous to the familiar case of classical instability, where it is completely natural to obtain a non-unitary coarse-grained evolution from the underlying unitary dynamics of the closed unstable system, with no need of restrictions or reformulations of the classical dynamical laws (see, e.g., Berkovitz, Frigg, and Kronz 2006). And, as in the classical case, there are many different coarse-grained states that can be defined on the same closed system, depending on which degrees of freedom are retained and which are disregarded.

- If the description given by the reduced state can be recovered from the perspective of the closed system as a coarse-grained description, it is not surprising that decoherence can be accounted for from a closed systems view. In fact, it is possible to study the loss of coherence in a system by considering the internal structure of the expectation values of its observables: decoherence is conceived in terms of the vanishing of the interference terms of the expectation values of certain observables of the closed system (Castagnino and Lombardi 2003, 2004, 2005a, 2005b). This closed systems view manifests the analogy between the classical statistical case and the quantum case (Fortin and Lombardi 2018): as Omnès (2001, 2002) has repeatedly claimed, decoherence is a coarse-grained phenomenon, a particular case of the phenomenon of irreversibility. Of course, this 'self-induced' approach to decoherence can also account for the results of the EID approach (Castagnino, Laura, and Lombardi 2007, Castagnino et al. 2008), but also dissolves the defining-systems problem: since the split of the closed system into an open subsystem and an environment is just a way of selecting the relevant observables among many possible decompositions, and since there is no privileged or 'essential' decomposition, then there is no need for an unequivocal criterion to decide where to place the cut between 'the' system and 'the' environment (Fortin and Lombardi 2016).
- From a closed-system perspective, the problem of locality no longer consists in accounting for correlations between the properties of non-interacting and possibly distant objects. Since the closed system is a single whole, EPR-correlations are correlations between properties of a single system, which are natural even in a classical ontological domain (see Holik et al. 2022: Section 3.2). In turn, the problem of indistinguishability no longer consists in explaining why a permutation of elemental particles of the same kind does not lead to a different complexion in quantum statistics. From this perspective, indistinguishability is not a relation between particles but turns out to be

an internal symmetry of the closed system conceived as an indivisible whole (see Holik et al. 2022: Section 3.3).

The top-down approach to ST was not only proposed from a general viewpoint, but was also embodied in a specific interpretation. The so-called Modal-Hamiltonian Interpretation (MHI), a member of the modal family (see Lombardi and Dieks 2021), endows the Hamiltonian of the closed system with a determining role (Lombardi and Castagnino 2008). According to the MHI's actualisation rule, any observable that does not commute with and/or does not have the symmetries of the Hamiltonian cannot acquire a definite actual value, since this actualisation would break the symmetry of the closed system in an arbitrary way. This rule has been applied to many physical situations—e.g., the free particle with spin, the harmonic oscillator, the free hydrogen atom, the Zeeman effect, the fine structure, and the Born–Oppenheimer approximation—resulting in descriptions consistent with empirical evidence (see Lombardi and Castagnino 2008: Section 5). The interpretation has also proved to be adequate in other situations, such as the problem of optical isomerism in chemistry (Fortin, Lombardi, and Martínez González 2018), and the emergence of a relational quantum event-time (Fortin, Lombardi, and Pasqualini 2022).

The MHI can also be formulated under a Galilean-invariant form, in terms of the Casimir operators of the Galilean group (Ardenghi, Castagnino, and Lombardi 2009, Lombardi, Castagnino, and Ardenghi 2010). Moreover, the MHI's actualisation rule can be transferred to the relativistic field by changing the symmetry group accordingly: the definite-valued observables are those represented by the Casimir operators of the Poincaré group, that is, the mass operator and the squared spin operator (Ardenghi, Castagnino, and Lombardi 2011, Lombardi and Fortin 2015). This agrees with the usual assumption in quantum field theory that elemental particles always have definite values of mass and spin, and those values are precisely what define the different kinds of elemental particles of the theory.

From an ontological viewpoint, the MHI proposes an ontology of properties, lacking the category of individual (da Costa, Lombardi, and Lastiri 2013, da Costa and Lombardi 2014), but not preventing the emergence of particles under particular circumstances (Lombardi and Dieks 2016). This ontological view also supplies an adequate answer to the problem of the entanglement of indistinguishable systems (Fortin and Lombardi 2022, Lombardi 2023).

3.4 Is a Closed Systems View Sustainable?

Michael Cuffaro and Stephan Hartmann (2021, 2024, this volume) offer a very detailed and articulated defense of the fundamentality of the open systems view of quantum mechanics, embodied in what they call *general quantum theory of open*

systems (GT). The top-down approach presented in the previous section is clearly at odds with GT. Therefore, this chapter cannot be closed without considering how fatal Cuffaro and Hartmann's arguments are for that closed systems view. In the space available I do not intend to analyse in detail such an extensive and well-argued work, but there are some points of the argumentation that I would like to comment on, because this task would make the authors' strong conclusions more flexible.

- *On compatibility*. An argument used to support the virtues of GT is based on the fact that other sciences conceive their systems as primarily open, so an open systems view in physics would promote the compatibility between sciences. However, this aim of unification needs to be taken with care. In other sciences, systems are *essentially* open, because their very existence depends on their exchanges with the environment: for instance, a biological system cannot be conceived as completely isolated, even in principle. By contrast, there is no theoretical reason that prevents conceiving a physical system as closed, because its existence does not depend on its openness. And this is mainly how physical systems have been treated throughout the history of physics.
- *On existence*. When considering how GT is to be interpreted, it is said: 'The open systems view does not, per se, deny that closed systems exist. It merely denies that only closed systems exist' (Cuffaro & Hartmann 2024: 17). Even if one accepts to discuss the issue in terms of existence, it is necessary to take into account that closed systems and open systems do not exist in the same way: whereas a closed system needs nothing else to exist, open systems only exist relatively to others. Given a quantum-closed system, there is no mereological way of identifying *the* open subsystems that compose it, given the many ways that it can be decomposed into parts. Given an open quantum system, by contrast, it cannot be said whether it exists, or its parts exist, or another open system exists of which it is a part: there is no univocal way to identify the open systems that are proposed as the fundamental objects of the open systems view. Perhaps this might be not an insurmountable obstacle to GT, but the relativity of open quantum systems cannot be overlooked in the defense of an open systems view.
- *On density operators*. Related to the previous point is the fact that this presentation of GT relies too much on the difference between state vectors and density operators. Although recognising the difference between proper and improper mixtures in the first pages, later it is said that density operators ρ describe 'systems that evolve, *in general*, non-unitarily' (italics mine), as if the case of unitary evolution were the particular case in which the state is pure and the density operator ρ can be expressed in terms of the state vector as $|\psi\rangle\langle\psi|$. But this ignores the just mentioned difference:

when the density operator represents a proper mixture, it evolves unitarily; it is the *reduced* density operator representing an improper mixture that evolves non-unitarily. The fact that proper and improper mixtures cannot be distinguished from a mathematical viewpoint is what leads to continuous confusions in the literature. However, this mathematical feature essentially depends on the formalism used to express the theory, in particular, the Hilbert space formalism; the situation might be different in other cases. For instance, it has been proved (Masillo, Scolarici, and Sozzo 2009) that proper and improper mixtures are represented by different density operators in the so-called quaternionic formulation of quantum mechanics (see Adler 1995); hence they can be distinguished also from a mathematical viewpoint. Moreover, this mathematical representation is compatible with their different time-evolutions as represented in the orthodox Hilbert space formalism.

- *On interpretation*. In the context of interpreting ST and GT, Cuffaro and Hartmann claim: 'if one takes ST to be a candidate fundamental theoretical framework, then there are only two families of interpretational options available: Everett and orthodox; and on either option one should conclude that ST is about open systems (at least in part) despite being formulated from the closed systems point of view.' In line with this claim, modal interpretations are not mentioned throughout the text. This may be due to a historical contingency. The traditional modal interpretations were specifically designed to solve the measurement problem, and they successfully achieved this goal in the case of ideal measurements. However, a series of articles in the 90s (Albert and Loewer 1990, 1991, 1993, Elby 1993, Ruetsche 1995) showed that those traditional approaches based on the modal view did not select the right properties for the apparatus in non-ideal measurements. As perfectly ideal measurements can never be obtained in practice, this shortcoming was considered a 'silver bullet' to kill modal interpretations (Harvey Brown, cited in Bacciagaluppi and Hemmo 1996). This explains, at least in part, the decline of the interest in modal interpretations since the end of the 90s. What was not sufficiently noted at that time was that the difficulties of those original modal interpretations in dealing with non-ideal measurements were due not to their modal nature, but to the fact that their rule of definite-value ascription made the set of definite-valued observables dependent on the instantaneous state of the system. An author who did realise this point was Jeffrey Bub, whose preference for Bohmian mechanics in those days can be understood in this context. In fact, if Bohmian mechanics is conceived as a member of the modal family, whose definite-valued observables are defined by the position observable (Bub 1997), it turns out to be a natural alternative given the difficulties of the original modal versions. However, what was not realised at that time is that position is not the only observable that can be appealed to in order to define the state-independent rule of definite-value ascription in a

modal interpretation. It is at this point that MHI enters the scene: it endows the Hamiltonian of the closed quantum system with the role of selecting the definite-valued observables. With this strategy, MHI not only solves the problems of the original modal interpretations, but can also be successfully applied to many physical situations. Unfortunately, this was still not enough to rehabilitate modal interpretations in the eyes of many philosophers of physics.

- *On laws*. In several parts of the text it is asserted that the universe is the only closed system that truly exists—and even this can be called into question given the existence of black holes: any other system is inevitably open, because of the influence of the gravitational field, the vacuum fluctuations, and other physical phenomena. It is even claimed that *real* systems are always open, and that the idea of closed system is a mere abstraction even according to orthodox interpretations formulated from the closed-system point of view. Let us concede that all this is true: does it mean that physics should be rewritten from the open systems view? Does it mean that we must set aside the Schrödinger—or the Liouville–von Neumann—equation in favour of the Lindblad equation in quantum mechanics? A positive answer to these questions implicitly assumes that the laws of physics describe 'real' systems in a direct way, that laws depict phenomena in the world by supplying their mirror images. However, this assumption has been repeatedly challenged by arguing that physical laws refer to highly idealised and in general non-existent situations (see Cartwright 1983 as the locus classicus). The relationship between laws and reality is always mediated by models, which incorporate approximations and idealisations in an inescapable way (see, e.g., Morgan and Morrison 1999). The validity of physical laws rests on the success of scientists in approximating real phenomena to these idealised situations by means of different strategies embodied in models. Newton's law of inertia claims that, if the net external force acting on a body is zero, the velocity of its center of mass remains constant. It may be the case, and it is quite plausible that this may indeed be the case, that there is no body in the universe such that the net external force acting on it is perfectly zero. Nevertheless, the law was never questioned for this reason to the extent that it is part of an empirically successful theory: this is the result of the so-called counterfactual force of laws. Michael Matthews (1987, 2004) tells the story of how the isochrony of the pendulum was vigorously resisted by Guidobaldo del Monte, Galileo's protector and one of the greatest experts in 17th-century mechanics. From his Aristotelian perspective, del Monte insisted that pendulums are, in fact, not isochronous: lighter and heavier pendulums do not have the same period, nor do those of different lengths; moreover, all pendulums eventually stop after a sufficient number of oscillations. Galileo replied that such results referred only to real pendulums, but that if one studies ideal

pendulums, where friction, air resistance, and the weight of the string do not exist, then they will certainly be found to be isochronous. Such an argument was never accepted by del Monte, who was convinced that physics should refer to the real world and not to a supposed ideal world without resistance and friction.

- *On practice*. When discussing the epistemic fundamentality of GT, it is said that, in practice, the epistemic order is beginning with the mixed state of the open system, and then assigning a pure state to it by throwing away some information. Of course, this may be the strategy in some cases. The question is if it is always the strategy when applying quantum mechanics in practice. Let us consider the paradigmatic case of the application of ST to molecules in chemistry: the first step is always the attempt to describe the isolated molecule, and even this task requires multiple approximations that increase their complexity when more precision is required and when molecules grow themselves in complexity. The difficulties in this field are high enough as to increase them by considering the influence of the environment from the beginning. It is this epistemic approach what leads to what Robin Hendry (2010), following Guy Woolley and Brian Sutcliffe (1977), calls 'the symmetry problem': since the interactions embodied in the Hamiltonian of the closed molecule are Coulombic, the solutions of the Schrödinger equation (the eigenstates of the resultant Hamiltonian) are rotationally symmetric; however, the rotational asymmetry of polyatomic molecules is essential in the explanation of their chemical behaviour (see also Sutcliffe and Woolley 2022). Some authors have proposed solutions to this and other problems related to the notion of molecular structure by considering the interaction of the molecule with its environment (see, e.g., Fortin, Lombardi, and Martínez González 2018, Seifert 2022), but this does not cancel the fact that the epistemic departing point is the molecule as a closed system.
- *On fundamentality*. It is true that Cuffaro and Hartmann's argumentation is not based on a reductive approach, in the traditional sense of the search for the 'fundamental' discipline and theory; on the contrary, the aim is to bring physics closer to special sciences. Nevertheless, their emphasis on fundamentality, mainly in its ontological meaning, seem to embody a metaphysically realist view according to which science tends to describe, or should describe the only 'real' and independent ontology. But this is not the only possible view about scientific knowledge. In particular, pluralist perspectives challenge the idea of the unity of science and stress the diversity inherent to science in its various aspects: theoretical, practical, and ontological (see, e.g., Ludwig and Ruphy 2021). I myself defend a Kantian-rooted pluralism, according to which ontology is partially constituted by the categorial-conceptual-practical framework underlying each field of scientific knowledge (Lombardi 2022). In any case, strong claims of fundamentality find no comfortable place in the context of scientific pluralism.

3.5 Final Remarks

Everything I said in the previous sections in no way intends to reject an open systems view of quantum mechanics, but only attempts to make it more flexible. Cuffaro and Hartmann's work offers a very interesting proposal that deserves to be put in dialogue with the more traditional ways of conceiving quantum mechanics.

Bibliography

Adler, S. L. (1995), *Quaternionic Quantum Mechanics and Quantum Fields* (Oxford: Oxford University Press).

Albert, D., and Loewer, B. (1990), 'Wanted Dead or Alive: Two Attempts to Solve Schrödinger's Paradox', in A. Fine, M. Forbes, and L. Wessels (eds.), *Proceedings of the 1990 Biennial Meeting of the Philosophy of Science Association, Vol. 1* (Philosophy of Science Association), 277–285.

Albert, D., and Loewer, B. (1991), 'The Measurement Problem: Some "Solutions"', in *Synthese* 86/1: 87–98.

Albert, D., and Loewer, B. (1993), 'Non-Ideal Measurements', in *Foundations of Physics Letters* 6/4: 297–305.

Ardenghi, J. S., Castagnino, M., and Lombardi, O. (2009), 'Quantum Mechanics: Modal Interpretation and Galilean Transformations' in *Foundations of Physics* 39/9: 1023–1045.

Bacciagaluppi, G., and Hemmo, M. (1996), 'Modal Interpretations, Decoherence and Measurements', in *Studies in History and Philosophy of Modern Physics* 27/3: 239–277.

Barnum, H., Knill, E., Ortiz, G., Somma, R., and Viola, L. (2003), 'Generalizations of Entanglement Based on Coherent States and Convex Sets', in *Physical Review A* 68/3: 032308.

Barnum, H., Knill, E., Ortiz, G., Somma, R., and Viola, L. (2004), 'A Subsystem-Independent Generalization of Entanglement', in *Physical Review Letters* 92/10: 107902.

Bellomo, G., Majtey, A. P., Plastino, A. R. and Plastino, A. (2014), 'Quantum Correlations from Classically Correlated States', in *Physica A* 405: 260–266.

Berkovitz, J., Frigg, R., and Kronz, F. (2006), 'The Ergodic Hierarchy, Randomness and Hamiltonian Chaos', in *Studies in History and Philosophy of Modern Physics* 37/4: 661–691.

Bub, J. (1997), *Interpreting the Quantum World* (Cambridge: Cambridge University Press).

Cartwright, N. (1983), *How the Laws of Physics Lie* (Oxford: Clarendon Press).

Castagnino, M., Fortin, S., Laura, R., and Lombardi, O. (2008), 'A General Theoretical Framework for Decoherence in Open and Closed Systems', in *Classical and Quantum Gravity* 25/15: 154002.

Castagnino, M., Fortin, S., and Lombardi, O. (2010), 'Is the Decoherence of a System the Result of its Interaction with the Environment?', in *Modern Physics Letters A* 25/17: 1431–1439.

Castagnino, M., Laura, R., and Lombardi, O. (2007), 'A General Conceptual Framework for Decoherence in Closed and Open Systems', in *Philosophy of Science* 74/5: 968–980.

Castagnino, M., and Lombardi, O. (2003), 'The Self-Induced Approach to Decoherence in Cosmology', in *International Journal of Theoretical Physics* 42/6: 1281–1299.

Castagnino, M., and Lombardi, O. (2004), 'Self-Induced Decoherence: A New Approach', in *Studies in History and Philosophy of Modern Physics* 35/1: 73–107.

Castagnino, M., and Lombardi, O. (2005a), 'Self-Induced Decoherence and the Classical Limit of Quantum Mechanics', in *Philosophy of Science* 72/5: 764–776.
Castagnino, M., and Lombardi, O. (2005b), 'Decoherence Time in Self-Induced Decoherence', in *Physical Review A* 72/1: 012102.
da Costa, N., and Lombardi, O. (2014), 'Quantum Mechanics: Ontology Without Individuals', in *Foundations of Physics* 44/12: 1246–1257.
da Costa, N., Lombardi, O., and Lastiri, M. (2013), 'A Modal Ontology of Properties for Quantum Mechanics', in *Synthese* 190/17: 3671–3693.
Cuffaro, M. E. and Hartmann, S. (2021), 'The Open Systems View', *arXiv preprint arXiv:2112.11095*.
Cuffaro, M. E., and S. Hartmann (2024), 'The Open Systems View', in *Philosophy of Physics* 2: 6.
Dugić, M., and Jeknić-Dugić, J. (2008), 'What is "System": The Information-Theoretic Arguments', in *International Journal of Theoretical Physics* 47/3: 805–813.
Earman, J. (2015), 'Some Puzzles and Unresolved Issues about Quantum Entanglement', in *Erkenntnis* 80/2: 303–337.
Elby, A. (1993), 'Why 'Modal' Interpretations Don't Solve the Measurement Problem', in *Foundations of Physics Letters* 6/1: 5–19.
Fortin, S., and Lombardi, O. (2014), 'Partial Traces in Decoherence and in Interpretation: What do Reduced States Refer To?', in *Foundations of Physics* 44/4: 426–446.
Fortin, S., and Lombardi, O. (2016), 'A Top-Down View of the Classical Limit of Quantum Mechanics', in R. E. Kastner, J. Jeknić-Dugić, and G. Jaroszkiewicz (eds.), *Quantum Structural Studies: Classical Emergence from the Quantum Level* (World Scientific), 435–468.
Fortin, S., and Lombardi, O. (2018), 'Understanding Decoherence as an Irreversible Process', in *International Journal of Quantum Foundations* 4: 247–267.
Fortin, S., and Lombardi, O. (2022), 'Entanglement and Indistinguishability in a Quantum Ontology of Properties', in *Studies in History and Philosophy of Science* 91: 234–243.
Fortin, S., Lombardi, O., and Martínez González, J. C. (2018), 'A New Application of the Modal-Hamiltonian Interpretation of Quantum Mechanics: The Problem of Optical Isomerism', in *Studies in History and Philosophy of Modern Physics* 62: 123–135.
Fortin, S., Lombardi, O., and Pasqualini, M. (2022), 'Relational Event-Time in Quantum Mechanics', in *Foundations of Physics* 52/1: 10.
Franco, M., and Calzetta, E. (2011), 'Decoherence in the Cosmic Background Radiation', in *Classical and Quantum Gravity* 28/14: 145024.
Harshman, N. L. (2012), 'Observables and Entanglement in the Two-Body System', in *AIP Conference Proceedings* 1508/1: 386–390.
Harshman, N. L., and Ranade, K. S. (2011), 'Observables Can be Tailored to Change the Entanglement of Any Pure State', in *Physical Review A* 84/1: 012303.
Harshman, N. L., and Wickramasekara, S. (2007), 'Galilean and Dynamical Invariance of Entanglement in Particle Scattering', in *Physical Review Letters* 98/8: 080406.
Harshman, N. L. (2016), 'Symmetry and Natural Quantum Structures for Three-Particles in One-Dimension', in R. E. Kastner, J. Jeknić-Dugić, and G. Jaroszkiewicz (eds.), *Quantum Structural Studies: Classical Emergence from the Quantum Level* (World Scientific), 373–400.
Healey, R., and Gomes, H. (2022), 'Holism and Nonseparability in Physics', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022 Edition), URL = https://plato.stanford.edu/archives/spr2022/entries/physics-holism/.

Hendry, R. F. (2008), 'Two Conceptions of the Chemical Bond', in *Philosophy of Science* 75/5: 909–920.
Holik, F., Krause, D., Jorge, J. P., and Lombardi, O. (2022), 'Quasi-Set Theory: A Formal Approach to a Quantum Ontology of Properties', in *Synthese* 200/5: 401.
Howard, D. (1985), 'Einstein on Locality and Separability', in *Studies in History and Philosophy of Science* 16/3: 171–201.
Kastner, R. E., Jeknić-Dugić, J., and Jaroszkiewicz, G. (eds.) (2016), *Quantum Structural Studies: Classical Emergence from the Quantum Level* (World Scientific).
Lee, H.-W., and Kim, J. (2000), 'Quantum Teleportation and Bell's Inequality Using Single-Particle Entanglement', in *Physical Review A* 63/1: 012305.
Lombardi, O. (2022), 'A Kantian-Rooted Pluralist Realism for Science', in C. Soto (ed.), *Current Debates in Philosophy of Science: In Honor of Roberto Torretti* (Springer Synthese Library), 77–95.
Lombardi, O. (2023), 'Entanglement and Indistinguishability: Facing Some Challenges from a New Perspective, in *Philosophical Transactions of the Royal Society A* 381/2255: 20220101.
Lombardi, O., and Castagnino, M. (2008), 'A Modal-Hamiltonian Interpretation of Quantum Mechanics', in *Studies in History and Philosophy of Modern Physics* 39/2: 380–443.
Lombardi. O., and Dieks, D. (2016), 'Particles in a Quantum Ontology of Properties', in T. Bigaj and C. Wüthrich (eds.), *Metaphysics in Contemporary Physics* (Brill-Rodopi), 123–143.
Lombardi, O., and Dieks, D. (2021), 'Modal Interpretations of Quantum Mechanics', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), URL = https://plato.stanford.edu/archives/win2021/entries/qm-modal/.
Lombardi, O., and Fortin, S. (2015), 'The Role of Symmetry in the Interpretation of Quantum Mechanics', in *Electronic Journal of Theoretical Physics* 12/34: 255–272.
Lombardi, O., Fortin, S., and Castagnino, M. (2012), 'The Problem of Identifying the System and the Environment in the Phenomenon of Decoherence', in H. W. de Regt, S. Hartmann, and S. Okasha (eds.), *Philosophical Issues in the Sciences Volume 3* (Springer), 161–174.
Lombardo, F., and Mazzitelli, D. (1996), 'Coarse Graining and Decoherence in Quantum Field Theory', in *Physical Review D* 53/4: 2001.
Ludwig, D., and Ruphy, S. (2021), 'Scientific pluralism', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), URL = https://plato.stanford.edu/archives/win2021/entries/scientific-pluralism/
Masillo, F., Scolarici, G., and Sozzo, S. (2009), 'Proper versus Improper Mixtures: Towards a Quaternionic Quantum Mechanics', in *Theoretical and Mathematical Physics* 160/1: 1006–1013.
Matthews, M. R. (1987), 'Galileo's Pendulum and the Objects of Science', in B. Arnstine and D. Arnstine (eds.), *Philosophy of Education* (Philosophy of Education Society), 309–319.
Matthews, M. R. (2004), 'Idealisation and Galileo's Pendulum Discoveries: Historical, Philosophical and Pedagogical Considerations', *Science & Education*, 13: 689–715.
Morgan, M., and Morrison, M. (eds.) (1999), *Models as Mediators* (Cambridge: Cambridge University Press).
Omnes, R. (2001), 'Decoherence: An Irreversible Process', in arXiv:quant-ph/0106006.
Omnes, R. (2002), 'Decoherence, Irreversibility and the Selection by Decoherence of Quantum States with Definite Probabilities', in *Physical Review A* 65/5: 052119.

Pan, J.-W., Bouwmeester, D., Weinfurter, H., and Zeilinger, A. (1998), 'Experimental Entanglement Swapping: Entangling Photons that Never Interacted', in *Physical Review Letters* 80/18: 3891–3894.

Ruetsche, L. (1995), 'Measurement Error and the Albert–Loewer Problem', in *Foundations of Physics Letters* 8/4: 327–344.

Seifert, V. A. (2022), 'Do Molecules have Structure in Isolation? How Models can Provide the Answer', in O. Lombardi, S. Fortin, and J. C. Martínez González (eds.), *Quantum Chemistry: Philosophical Perspectives in Modern Chemistry* (Springer Synthese Library), 125–143.

Sutcliffe, B. T., and Woolley, R. G. (2022), 'Is Chemistry Really Founded in Quantum Mechanics?', in O. Lombardi, S. Fortin, and J. C. Martínez González (eds.), *Quantum Chemistry: Philosophical Perspectives in Modern Chemistry* (Springer Synthese Library), 173–202.

Terra Cunha, M. O., Dunningham, J. A., and Vedral, V. (2007), 'Entanglement in Single-Particle Systems', in *Proceedings of the Royal Society A* 463/2085: 2277–2286.

Viola, L., and Barnum, H. (2010), 'Entanglement and Subsystems, Entanglement Beyond Subsystems, and All That', in A. Bokulich, and G. Jaeger (eds.), *Philosophy of Quantum Information and Entanglement* (Cambridge: Cambridge University Press), 16–43.

Woolley, R. G., and Sutcliffe, B. T. (1977), 'Molecular Structure and the Born–Oppenheimer Approximation', in *Chemical Physics Letters* 45/2: 393–398.

Zanardi, P. (2001), 'Virtual Quantum Systems', in *Physical Review Letters* 87/7: 077901.

Żukowski, M., Zeilinger, A., Horne, M. A., and Ekert, A. K. (1993), '"Event-Ready-Detectors" Bell Experiment via Entanglement Swapping', in *Physical Review Letters* 71/26: 4287–4290.

Zurek, W. H. (1982), 'Environment-Induced Superselection Rules', in *Physical Review D* 26/8: 1862–1880.

Zurek, W. H. (2000), 'Decoherence and Einselection', in P. Blanchard, D. Giulini, E. Joos, C. Kiefer, and I.-O. Stamatescu (eds.), *Decoherence: Theoretical, Experimental, and Conceptual Problems* (Springer-Verlag), 309–342

Zurek, W. H. (2003), 'Decoherence, Einselection, and the Quantum Origins of the Classical', in *Reviews of Modern Physics* 75/3: 715–776.

4
Quantum Systems Other than the Universe

David Wallace

4.1 Introduction: Physics as the Study of Autonomous Systems

Quantum theory is often described as our most fundamental theory of physics, and as such, the task of understanding what quantum theory tells us about the world is often taken to be the task of understanding what the world would be like if quantum theory in fact described it completely. To understand quantum theory in this way commits us (here I follow Wallace 2022a) to two assumptions: the *Fundamentality Assumption*, that the theory is *exactly true*, and the *Cosmological Assumption*, that the theory describes *the whole of reality*.[1] Of course, no-one literally thinks that any particular quantum theory—or indeed any theory we currently possess—actually satisfies either assumption, but still: much, perhaps most, interpretative work on quantum theory and indeed on theories of physics in general takes them as a starting point.

At the risk of cliché (cf. Quine 1951), we might call these two *dogmas* of interpretation—dogmas that, once made explicit, are very hard to argue for. A moment's observation of physics shows that hardly ever do physicists seek to model the whole universe, and hardly ever do they seek to model every last degree of freedom of even a finite system. What actually happens in physics, in almost all cases, is that a physical theory is used to model some subsystem of the universe, and to model that subsystem only on certain scales. (Even in general relativity, where our models are called 'spacetimes' and it is only natural to think of them as modelling entire universes, in practice the vast majority of the time we are actually using them to model isolated subsystems of a larger universe. Even when applied to cosmology, our models usually describe patches of a larger cosmos, and normally track only large-scale features of the geometry and matter distribution.)

Of course, it remains logically possible that our understanding of physics modelling practice is somehow derivative on the interpretation of theories under the Fundamentality and Cosmological Assumptions. But there seems little positive reason so to interpret them: what we actually have evidence for is the efficacy of physics in modelling scale-dependent features of subsystems, and it seems far more

David Wallace, *Quantum Systems Other than the Universe*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press. © Oxford University Press (2026).
DOI: 10.1093/9780198929277.003.0005

natural to look for a direct understanding of physical theories—and, in particular, quantum theories—as they are so used.

But *which* subsystems are the proper concern of physics? A very natural answer—one I (too quickly) endorsed in Wallace (2022a)—is that physics models *isolated* systems, systems that do not interact with any other system. But this has multiple problems. For one thing, it is not as straightforward as it might seem to say carefully what 'isolated' actually means here — the temptation is to say something like 'evolving the same way as it would if it were alone in the universe', but that is scarcely a counterfactual that we can assess empirically. For another, insofar as we can understand 'isolation' it seems to apply to systems that are dynamically isolated from one another in some (usually spatial) way, and does not naturally apply when we are modelling some degrees of freedom of a system and neglecting other, perhaps smaller-scale, degrees of freedom of that same system.

We could go formal. Quantum mechanics—at least as it is normally formulated—is a theory with unitary dynamics, where a system evolves under the Schrödinger equation with respect to some fixed Hamiltonian. We could *define* an isolated quantum system as a system that could be so modelled.

One problem with this definition is that unitarity does not coincide with isolation in the physical sense. Sometimes an environment interacts with a system not by introducing non-unitary dynamics but by adjusting—*renormalising*—the Hamiltonian of the system. But there is a much more serious problem, which also afflicts any attempt to make precise a more physical notion of isolation. The problem is that 'isolated system' under any reasonable disambiguation is inadequate to capture all the systems that physics (including quantum physics) in fact succeeds in modelling. The excited atom that decays through the emission of radiation; the pollen grain undergoing Brownian motion; the friction-bound system that eventually slows to a stop … all these are paradigmatically important, yet paradigmatically *non*-isolated, systems. Many such cases, indeed, are paradigmatically important *quantum* systems, but they are not isolated in either the substantive sense that they can be physically understood without consideration of their environment, or in the formal sense that their dynamics can be reduced to the Schrödinger evolution of pure states.

Let me suggest a better answer to the question of physics' proper concern. What physics models is *autonomous* systems, and a system is autonomous if it can be effectively modelled *as* a system with a state space and a dynamics, such that many initial states of the system are possible and that the dynamics tells us what the future states are as a (possibly stochastic) function of the present and past states. In particular, a system is an autonomous *quantum* system if it can be effectively modelled as a system with a Hilbert space and a dynamics for (not necessarily pure) quantum states on that Hilbert space.

All the systems I mention above are autonomous in this sense—of course various things must hold of the environment for the system to be autonomous, and

broad features of the environment (the temperature of the fluid in which the pollen grain moves, for instance) determine the dynamical parameters, but nonetheless it is possible to state, and test, a dynamical equation for the system's degrees of freedom, without regard for the *details* of the state of the environment.

Autonomy is neither trivial nor automatic: most subsystems of a complex interacting system will *not* be autonomous. (For instance, consider three bodies of similar masses interacting gravitationally: the three bodies form a complex interacting system and no two bodies can be modelled effectively, even stochastically, without tracking the third.) Nor need science be concerned only with autonomous systems. A non-autonomous system can be thought of as requiring not only initial-time data but a constant stream of data describing the influence of the environment, data which cannot itself be modelled (at least not as part of studying this system) but is simply an input. The human body, or an organ of the body, can be profitably studied this way: we want to know not what it does in isolation but what it does as a function of the way the environment impinges upon it over time.

But for the most part, *physics* is concerned with autonomous systems. The primary subject matter of physics is dynamics, and non-autonomous systems have no dynamics: their behaviour cannot be analysed dynamically except as a shadow on the wall, cast by the dynamics of a larger system.

My main goal in this chapter is to develop what an account of quantum systems looks like given that it should be a theory of *autonomous* systems, but not necessarily of isolated systems—still less systems that aim to model the universe. The starting point in this account (developed in section 4.2) is the observation that *formal* features of quantum theory—when combined with appropriate choices of system and dynamics—imply a robust notion of autonomy that is equivalent neither to unitary dynamics nor to the physical notion of isolation. We can obtain that notion by (i) writing a formally exact equation for the evolution of the reduced state of any subsystem of a unitarily evolving quantum system, and (ii) identifying physical assumptions under which that formally exact equation reduces (in some appropriate idealisation) to an autonomous dynamical equation for the subsystem alone. This analysis shows that in some sense it is inconsistent to suppose that quantum systems always and only obey unitary dynamics: the existence of large systems with unitary dynamics often entails the existence of smaller autonomous subsystems with non-unitary dynamics.

This observation, especially once we reject the Fundamentality and Cosmological Assumptions, might suggest that unitarity is a ladder that can be kicked away, and that we should replace it entirely with a generalised, non-unitary (perhaps also non-Markovian) notion of quantum dynamics. I do not think this is true. Quantum physics may not be about exactly-true models of the whole universe, but it does provide very powerful (essentially inductive) evidence for a thesis of *quantum reductionism*: that non-unitarity can and should always be understood

as the restriction to a subsystem of larger-scale unitary dynamics, and that larger-scale, phenomenological quantum physics is ultimately explicable microscopically in terms of specific, largely known, unitary quantum theories: non-relativistic particle mechanics and various quantum field theories. I develop this position in section 4.3, where I suggest four assumptions of quantum reductionism which in a way constitute a naturalistic replacement for the Fundamentality and Cosmological Assumptions.

I am not the first to develop a view of quantum systems along these general lines, and indeed this paper is substantially inspired by recent pioneering work by Cuffaro and Hartmann (2024). Their *Open Systems View* rejects the primacy of unitary dynamics and aims at analysing open quantum systems, not just isolated ones, and my second goal in this paper is to engage with their view and see how it relates to mine. In section 4.4 I argue that their view tacitly requires the notion of autonomy which I have described above, and that once this is recognised, the main difference between our views concerns the status of quantum reductionism: at least on one interpretation, the Open Systems View is committed to the rejection of most or all of my four assumptions of quantum reductionism. As I go on to argue, in large part I think this feature of their view needs to be rejected, for a mixture of reasons: partly that it makes the very notion of 'interaction' mysterious in a way which inhibits physics, partly because of the evidential support for reductionism in physics practice. But in section 4.5 I identify an important—if partial—exception to my arguments, concerned with how *irreversibility* enters into the dynamics of autonomous systems, and relate the differences between my view and the Open Systems View to unresolved issues in the philosophy of statistical mechanics.

The physics I discuss is fairly standard and for the most part I do not provide original references. An up-to-date standard monograph on open quantum systems is Breuer and Petruccione (2007).

4.2 Open Quantum Systems: Deriving Autonomy from Unitarity

Suppose that some unitarily evolving quantum system with Hilbert space $\mathcal{H}$ and Hamiltonian $\widehat{H}$ has a tensor-product decomposition into subsystems S and E, $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_E$, and a concomitant decomposition of its Hamiltonian,

$$\widehat{H} = \widehat{H}_S \otimes \widehat{1} + \widehat{1} \otimes \widehat{H}_E + \sum_n \widehat{C}_n \otimes \widehat{D}_n. \tag{1}$$

If the third term in this expression vanishes, $\mathcal{H}_S$ and $\mathcal{H}_E$ will define closed systems with Hamiltonians $\widehat{H}_S$, $\widehat{H}_E$ respectively. Even in this case, of course, the embedding of these systems in a larger system means that their joint state might be entangled, so that the appropriate notion of *state* for each subsystem is a mixed rather than

pure state. (This already suffices to demonstrate that once we drop the Cosmological Assumption, the principle that a quantum system must be in a pure state becomes untenable; cf. Wallace (2025) for further development of this observation in the context of quantum statistical mechanics.)

However, the vanishing of the interaction term is not the most general circumstance in which a subsystem of a larger quantum system is autonomous. To see this,[2] let S_E be a fixed (in general mixed) state of $\mathcal{H}_E$, which we will require to be invariant under the self-Hamiltonian $\widehat{H}_E$ of $\mathcal{H}_E$, $[S_E, \widehat{H}_E] = 0$. Then define the following superoperator $\mathbf{J}$ on the joint density operators of $\mathcal{H}_S \otimes \mathcal{H}_E$:

$$\mathbf{J}\rho = Tr_E\rho \otimes S_E \tag{2}$$

where Tr_E is the partial trace over $\mathcal{H}_E$, i. e., $Tr_E\rho$ is the reduced density operator of $\mathcal{H}_S$. Note that $\mathbf{J}$ is a projection, $\mathbf{J}^2 = \mathbf{J}$, projecting onto the subspace of states of form $\rho = \rho_S \otimes S_E$. (Also note that here and after, I (i) use **boldface** to label superoperators that act on the space of Hilbert-space operators; (ii) drop the 'hats' on mixed quantum states, to avoid cluttering the notation.)

Following Nakajima (1958) and Zwanzig (1961) we can now extract an integro-differential equation for ρ_S, the *Nakajima–Zwanzig equation*:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\rho_S(t) = -i\left[\widehat{H}_R, \rho_S(t)\right] + Tr_E \int_0^t \mathrm{d}\tau\, \mathbf{L}_H\mathbf{V}(\tau)(\mathbf{1}-\mathbf{J})\mathbf{L}_H(\rho_S(t-\tau)\otimes S_E) \\ + Tr_E\mathbf{L}_H\mathbf{V}(t)(\rho(0) - \rho_S(0)\otimes S_E)\end{aligned} \tag{3}$$

where:

- $\widehat{H}_R$ is the *renormalised system Hamiltonian*,

$$\widehat{H}_R = \widehat{H}_S + \sum_n \widehat{C}_n Tr(\widehat{D}_n S_E). \tag{4}$$

- $\mathbf{L}_H$ is the superoperator representing the adjoint action of the Hamiltonian,

$$\mathbf{L}_H\rho = -i\left[\widehat{H}, \rho\right]. \tag{5}$$

- $\mathbf{V}(t)$ is a modified propagator,

$$\mathbf{V}(t) = \mathrm{e}^{(\mathbf{1}-\mathbf{J})\mathbf{L}_H t}. \tag{6}$$

Equation (3) is exact, and so by itself tells us nothing about whether subsystem S has autonomous dynamics. However, in many cases of physical interest there is *separation of scales*, so that the timescale T_E on which the degrees of freedom

of E evolve—and, in particular, effectively equilibrate—is much faster than the timescale T_S on which the degrees of freedom of S evolve and interact with E. When this occurs (something that needs to be argued for in any specific case), then if we take S_E to be the appropriate equilibrium state (i.e., $S_E \propto \exp(-\beta \widehat{H}_E)$ for some inverse temperature β):

1. Since effective equilibration of a system means that the matrix elements of those operators describing collective degrees of freedom—such as those describing coupling with other systems—approach their equilibrium values, for $t \gg T_E$ we ought to be able to approximate ρ_0 by $\rho_S \otimes S_E$ in the third term, so that it approximately vanishes.
2. Since the second term tracks the degree to which previous states of S leave a memory trace in E, and since such traces ought to be effectively unravelled by equilibration, the kernel in that second term ought to fall off on timescales $\sim T_E$, so that for $t \gg T_E$ we can replace the upper limit t of integration with some fixed large (but sub-recurrent) time T.

Equation (3) then simplifies to

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_S(t) = -i\left[\widehat{H}_R, \rho_S(t)\right] + Tr_E \int_0^T \mathrm{d}\tau\, \mathbf{K}(\tau)(\rho_S(t-\tau) \otimes S_E) \tag{7}$$

where $\mathbf{K}(\tau) = \mathbf{J}\mathbf{L}_H\mathbf{V}(\tau)(\mathbf{1} - \mathbf{J})\mathbf{L}_H$. This is an autonomous, time-translation-invariant, evolution equation for ρ_S. It is, however, *non-Markovian*: the system has memory, and its future evolution depends on its past and not just its present state.

Systems which are *sufficiently* non-Markovian are often autonomous only by courtesy. The interactions with the system's environment leave traces in the system's past state, and those traces can be used to reconstruct the environment state. (It is plausible that one could write down a wildly non-Markovian 'autonomous' dynamics for one body in the three-body problem.) Autonomy requires that the kernel $\mathbf{K}(\tau)$ falls off quickly compared to the timescales on which the system evolves, so that only its *recent* past is dynamically relevant. We have already assumed this, however, and so equation (7) indeed describes an autonomous dynamics.

In many cases we can further assume that the memory timescale is short enough that $\rho_S(t)$ is effectively static on that timescale, in which case the second term can be approximated as

$$Tr_E \int_0^T \mathrm{d}\tau\, \mathbf{K}(\tau)\rho_S(t-\tau) \otimes S_E \simeq Tr_E \int_0^T \mathrm{d}\tau\, \mathbf{K}(\tau)(\rho_S(t) \otimes S_E). \tag{8}$$

If we then define $\mathbf{\Lambda}$ by

$$\mathbf{\Lambda}\rho_S = Tr_E \int_0^T \mathrm{d}\tau\, \mathbf{K}(\tau)(\rho_S \otimes S_E) \tag{9}$$

we have extracted a *Markovian*, but in general non-unitary, dynamical equation for ρ_S:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_S(t) = -i\left[\widehat{H}_R, \rho_S(t)\right] + \mathbf{\Lambda}\rho_S(t). \tag{10}$$

Fairly general results from the theory of open quantum systems (see, e.g., Nielsen and Chuang 2000) tell us that in this case, $\mathbf{\Lambda}$ must be expressible in Lindblad form, that is, there must exist operators $\widehat{L}_i$ such that

$$\mathbf{\Lambda}\rho = \sum_i \left(\widehat{L}_i \rho \widehat{L}_i^\dagger + \frac{1}{2}\left(\widehat{L}_i \widehat{L}_i^\dagger \rho + \rho \widehat{L}_i \widehat{L}_i^\dagger\right)\right). \tag{11}$$

Finally, in some cases the $\mathbf{\Lambda}$ term in (7) can be neglected entirely—for instance, it might be that the energies required to excite the environment out of its ground state are too large compared to the characteristic system energies. In this case the system's evolution remains unitary; however, it is governed by the renormalised Hamiltonian $\widehat{H}_R$, not by the 'bare' self-Hamiltonian $\widehat{H}_S$, and so the influence of the environment is not entirely absent.

To sum up, we have a hierarchy of possibilities for a quantum system. At the bottom of the hierarchy, the system might be a 'system' only by courtesy, inextricably dynamically interwoven with its environment and studyable only in conjunction with it. Alternatively, it might be autonomous, governed by a dynamics which does not depend on the fine details of the environment (in the case we have considered, it depends only on the temperature) and which, if it has memory effects, has those effects only on timescales fairly short compared to its dynamical timescale. Above that in the hierarchy, it might be Markovian, with the memory effects idealisable into an instantaneous dynamics. And if those dissipative effects are entirely absent or can be idealised as such, the system is unitary.

Note that (as anticipated in section 4.1) 'isolated' does not appear in this hierarchy, and indeed is not straightforwardly definable in formal terms. It cannot be identified with 'unitary': as we have seen, even if dissipative effects of other systems can be neglected, other systems' influence can appear as a renormalisation. And this is common in physics: the famous renormalisations of quantum field theory have exactly this form, for instance, where the effects of high-energy physics manifest at lower energies not through any violation of unitarity, but through the renormalisation of the parameters in the unitary dynamics. (See Boyanovsky 2015 for an explicit presentation of quantum-field-theoretic renormalisation in this form.)

4.3 Four Dogmas of Quantum Reductionism

We have seen that it is not self-consistent to suppose that quantum systems invariably or even usually have unitary dynamics. Once we realise that quantum physics, like physics generally, is concerned with autonomous subsystems of the universe and not simply with the universe as a whole or those of its subsystems that can be treated as entirely isolated, we have to recognise that quantum systems generally must be allowed to have mixed states and that quantum dynamics need not be Markovian and, even when Markovian, need not be unitary.

Once we recognise this, we could in principle adopt 'quantum pluralism': the belief that there is nothing particularly special about unitary dynamics (or pure states) and that the right way to do quantum mechanics is simply to discover, system by system, what that system's dynamics actually are, free of any prejudice that they must be unitary. Quantum pluralism is a fundamentally disunified quantum metaphysics, very reminiscent of Nancy Cartwright's 'dappled world' metaphysics, in which we abandon any pretence for universality of any dynamical laws and work system-by-system, and quite closely related to Cuffaro and Hartmann's 'Open Systems View' (to which I will return in section 4.4).

But the vast majority of practicing physicists do not practice quantum pluralism, and I think they are correct not to do so. Physics practice is overwhelmingly committed to what I call *quantum reductionism*, a position that is encapsulated by four principles—four dogmas of quantum reductionism, if you like. The principles can be understood both methodologically (this is how, as practicing physicists, we *should* explore the physics of quantum systems) and metaphysically (this is, in fact, the *truth* about quantum systems). I should be clear that I use 'dogma' playfully here: in my view the four 'dogmas' of quantum reductionism are fully justified methodologically and probably largely correct metaphysically, and indeed I offer them as a replacement for the two dogmas we have already encountered (the Fundamentality and Cosmological Assumptions).

The first principle is that *non-unitarity implies interaction*. A unitarily evolving quantum system may or may not require us to consider its interactions with other systems in order to explain its dynamics (recall renormalisation), but if a system is non-unitarily evolving it is always doing so because of its interactions with other systems—or at least, denying this is an absolute last resort. More specifically: the presence of memory terms always indicates that some additional and unmodelled degrees of freedom are interacting with the system; even in the absence of memory terms, non-unitarity means that the system is becoming entangled with some other system. Methodologically, this principle advises us to look first for interaction-based explanations of non-unitarity, and posit intrinsic non-unitarity only as a last resort; metaphysically, it says that *in fact* non-unitarity always has an external cause.

The second principle is that *interaction is derived from dilation*. By a *dilation* of an autonomous quantum system, I mean the inclusion of that system in a

larger autonomous quantum system, such that the dynamics of the original system can be derived from the dynamics of the larger system by the means sketched in section 4.2. Methodologically, this principle advises us to look for a dilation-based account of interaction; metaphysically, it says that such an account exists even if we haven't found it yet.

Section 4.2 made only minor use of the unitarity of the larger system. The superoperator $\mathbf{L}_H$ defined the dynamics of that system, and I took $\mathbf{L}_H\rho = -i[\widehat{H}, \rho]$, but the derivation of equation (3) would go through largely unchanged if $\mathbf{L}_H$ were itself a non-unitary dynamics. (The main difference is that the first term in the expression would become non-unitary.) But in practice this is not how dilations are normally done in physics, and this is the third principle: that *the dilations that explain interaction are to larger unitary systems*. More precisely: they are to systems that can be idealised as unitary at least for the purposes of explaining the subsystem's dynamics: perhaps their own dynamics, modelled at a sufficiently high level of precision, requires non-unitary terms, but those terms make no significant contribution to the subsystem's autonomous dynamics. (If they did, then by the third principle we did not dilate far enough—but the third principle commits us not just to some in-principle possibility of dilating to a unitary system, but to the practical strategy of modelling interaction that way.) Again, methodogically speaking this is advice to the modeller; metaphysically speaking it is a substantive claim about the detailed origin of interaction terms.

At a purely formal level, the dilational assumptions in the second and third principle are almost empty: Stinespring's 1955 dilation theorem (see, e.g., Nielsen and Chuang 2000 for a modern presentation) says that (under weak additional assumptions; cf. Cuffaro and Hartmann 2024: section 2.2.2) any well-behaved linear dynamics for a quantum system is the restriction to that system of *some* unitary dynamics on *some* subsystem. The second and third assumptions get their bite from the requirement that the dilation needs to be not just to a fictional system but to some concretely specified system whose dynamics can themselves actually be observed and measured.

Indeed, in physics practice it is ultimately insufficient merely to posit the dynamics of that larger system, even if those posits are observationally confirmed. The fourth principle is that the dynamics of any autonomous quantum system should ultimately be derived, via restriction to subsystems and/or the averaging over microscopic degrees of freedom (insofar as these are distinct), and under appropriate restrictions on a system's state (e.g., fixing temperature or a regime of symmetry breaking), from a short list of 'microscopic'—and invariably unitary—quantum theories. The main entries on the list are:

1. Nonrelativistic quantum mechanics: The nonrelativistic physics of nuclei and electrons, interacting via Coulomb coupling, occasionally also by Newtonian gravitational coupling, and by coupling to background electric, magnetic, and gravitatational fields.

2. Relativistic quantum mechanics: The relativistic generalisation of (1), under the self-consistency assumption that interparticle interactions are gentle enough not to lead to particle creation.
3. Quantum optics, the hybrid theory of the quantised electromagnetic field coupled to nonrelativistic nuclei and electrons.
4. Quantum electrodynamics, describing relativistic electrons and nuclei (the latter treated as pointlike) interacting with the quantised electromagnetic field.
5. The Standard Model of Particle Physics, along with its various effective-field-theory approximations (which in fact include (1)–(4)).
6. Low-energy quantum gravity, the effective field theory of gravitation coupled to quantum matter. (Strictly speaking this is the only item we need on the list: (5), and hence (1)–(4), are derivable from it.)

Methodologically, the fourth principle advises us to look to theories on this short list when constructing and/or justifying phenomenological models; metaphysically, it is the reductionist thesis that in fact the systems we encounter are ultimately governed by the theories on the list.

The fourth principle is not inviolate, and is routinely violated in places where 'new physics' is expected or at least not ruled out: examples include dark matter, beyond-Standard-Model particle physics, and the observationally underconstrained mechanisms of CP violation and neutrino mass. But the positing of new physics (like fundamental non-unitarity) is normally a last resort, and this is especially true in terrestrial, low-energy situations—positing a new particle to explain why some high-temperature superconductor works, for instance, would require extremely powerful evidence and extensive argument as to why that particle had not shown up in other contexts.

Why think these four principles in fact constitute standard practice in physics? A really systematic demonstration would require quantitative methods well beyond the scope of this paper, but I will motivate it through two examples. The first and most important is non-equilibrium quantum statistical mechanics, where pretty much any derivation I know adheres to them. The name of the game is to derive equations—complete with quantitatively determined parameters—for collective degrees of freedom such as the spatially averaged density of a fluid or a particle interacting with an environment. Invariably the derivation begins with a larger system whose dynamics are taken to be unitary; often those dynamics are taken as determined by the specific forms found in microphysical theories of nonrelativistic quantum mechanics or quantum optics (or, for more exotic applications, relativistic field theory), and where they are not, we are given at least qualitative reasons to think their general form is consistent with those microphysical theories (or else the theory is specifically described as a toy model, intended to illustrate some general principle rather than describe some specific system).

The distinction between 'phenomenological' and 'micro-based' equations, central in statistical mechanics, is exactly the distinction between equations posited simply to explain the data and equations derived from known microphysics—both have their place, but ultimately we aim to see why the former are a consequence of the latter.

The second is the empirical search for fundamental non-unitarity (as occurs in dynamical collapse theories like the GRW theory, continuous state localisation, or the Diosi–Penrose model; see, e.g., Bassi et al. 2023 for an up-to-date review of extant collapse theories and their empirical status). Here it is essential (the first principle) to rule out the possibility that the non-unitarity is the consequence of interaction with an environment: only if this occurs do we have a genuine violation of quantum mechanics and not just emergent non-unitarity. But (the second principle) it is constitutive of 'interaction with an environment' that the interaction can actually be modelled as the consequence of dynamical coupling with a second system. And (the third and fourth principles) evidence of fundamental non-unitarity does not require the experimenter to rule out *every conceivable* dynamical interaction with another system (that would be in practice impossible) but only to rule out those dynamical interactions that our microphysical theories give us reason to expect to be present.

(Of course, fundamental non-unitarity is new physics, and once we are considering new physics, other possibilities appear: perhaps the apparent non-unitarity is after all a consequence of entanglement with a hitherto-unsuspected new particle or field. No single experiment would suffice to falsify the whole quantum framework: the point is that one way or another such a discovery would point to new physics, but it would do so only because of the methodological commitment to attempt first to explain the observations in terms of known—and unitary—microphysical theories.)

Put aside whether the four principles correctly describe physics practice. Why think that they are correct? The obvious answer is just that the strategy of accounting for the dynamics of quantum systems in this way has been spectacularly successful. We have extremely good evidence that, for instance:

- where nuclear interactions and electromagnetic radiation can be neglected, systems are accurately described by nonrelativistic quantum mechanics;
- where nuclear interactions but not electromagnetic radiation can be neglected, systems are accurately described by quantum electrodynamics (and, in the nonrelativistic regime, by quantum optics);
- where effects at energy scales above 10 TeV can be neglected, systems are accurately described by the Standard model;
- where Planck-scale physics can be neglected, self-gravitating systems (insofar as quantum effects are also present) are accurately described by low-energy quantum gravity (cf. Wallace 2022b and references therein).

And in each of these cases, the 'can be neglected' clause is not empty tautology but can be quantitatively assessed (for instance, in terms of typical energy levels) and checked if necessary by evaluating leading-order corrections from a more fundamental description. In short, both the metaphysical and methodological versions of the four dogmas of quantum reductionism are justified by the evidence that we actually do know what the correct microphysical theories of the world are, outside some fairly exotic contexts, and that *in fact* those theories are unitary quantum theories.

Furthermore, this 'extremely good evidence' does not simply come from microphysical explorations: it does not, that is, just consist of evidence from particle accelerators and the like combined with an *a priori* commitment to reductionism. It consists in large part of the thousands of successes of the overlapping fields of (equilibrium and non-equilibrium) quantum statistical mechanics, condensed matter physics, and effective field theory in deriving empirically confirmed claims at the phenomenological level from quantum microphysics. We have not by any means explained *every* quantitatively known quantum dynamical system in this way, but we have explained a great many: to pick just a few examples, we have quantitatively confirmed accounts of

- the linear responses of many forms of matter to external perturbations (through the relation of linear-response-theory coefficients to two-time expectation values and the evaluation of the latter in perturbation theory; see Zwanzig 1965 for a classic review);
- the master equations governing atomic and nuclear decay of excited states;
- the phenomena of superconductivity and superfluidity in a wide class of solids and fluids via the spontaneous breaking of (respectively) local and global phase symmetry;
- the effective theory of pion scattering (via the identification of the pion with the Nambu–Goldstone boson associated to spontaneous breaking of chiral symmetry in quantum chromodynamics);
- the suppression of quantum interference in mesoscopic systems through decoherence theory.

Reductionist claims of this modest kind are fairly uncontroversial in physics: Anderson, for instance, is often held up by philosophers as an arch-critic of reduction, but his famous 1972 discussion of the subject begins thus:

> The reductionist hypothesis may still be a topic for controversy among philosophers, but among the great majority of active scientists I think it is accepted without question. The workings of our minds and bodies, and all the animate or inanimate matter of which we have any detailed knowledge, are assumed to

> be governed by the same set of fundamental laws, which except under certain extreme conditions we feel we know pretty well.

But Anderson isn't wrong that they are controversial in philosophy of science. And notwithstanding the various quantitative successes of quantum statistical mechanics, both philosophers and physicists have claimed that various paradoxes of principle block its derivations: the measurement problem, the thermodynamic limit, the irreversibility problem of statistical mechanics. (I return to the latter in section 4.5.) Defending the thesis systematically is far more than I can do here; I will settle for the weaker claims that (a) a methodological commitment to reductionism in this sense has been intimately entangled with quantum theory itself since its inception (and indeed before), and (b) that this methodology has been extremely successful scientifically.

Although my focus here is on quantum theory, modern physics' use of *classical* physics adheres to a modified version of the same four dogmas. Here too there is a plurality of dynamics, with many classical systems obeying stochastic dynamics (like the Langevin equation) and/or dissipative dynamics (like the Navier–Stokes equation) instead of Hamiltonian dynamics (the classical analogue of unitary dynamics). But physics is methodologically committed (1) to explaining non-Hamiltonian dynamics by dilation to larger systems, insofar as it can be explained classically at all, and (2) to model those larger systems via Hamiltonian dynamics. The difference is that in classical physics there is always the possibility of a *quantum* explanation of the form of a classical equation, and there is no simple, short list of classical microphysical theories that suffice to explain all of classical dynamics: unification to that degree requires quantum mechanics. For more on this point, see Wilson (2006: 196–7) and Wallace (2020a: section 4).

4.4 The Open Systems View: A Comparison

Cuffaro and Hartmann (2024; henceforth CH) have recently expounded and defended what they call the 'open systems view' of quantum mechanics, according to which we should take open, rather than closed, quantum systems as fundamental and treat closed systems as a special case of open systems.[3] The Open Systems View clearly has a lot in common with the view I have been developing and advocating; in this section I want to consider how the two are related.

First: what exactly is the Open Systems View? CH give two characterisations, which are somewhat subtly related. The first is a *physical characterisation*: an open system is one which is dynamically coupled to its environment, while a closed system is isolated from it. (Open systems are 'systems evolving under the

influence of their environment'.) The second is a *formal characterisation*: an open quantum system is a quantum system with dynamics that is linear and preserves the consistency of the Born probability rule but need not be unitary; a closed quantum system has unitary dynamics. (CH are mostly interested in Markovian open systems, but I don't think the Markovian restriction is essential to them.)

Thus stated, the Open Systems View seems compatible with the present view, but it contains an important tacit assumption worth making explicit: the 'open systems' CH consider are required to be *autonomous* systems. In general, as we have seen, 'systems evolving under the influence of their environment' might have no self-contained dynamics at all, but CH are specifically concerned with systems that do have well-defined dynamics; they simply want to drop the assumption of unitarity.

Once this is recognised, one of CH's original arguments for the Open Systems View loses its force: they argue that 'there are reasons stemming from physics to believe that the cosmos is the *only* closed system that truly exists', because no system other than (perhaps) the cosmos can be entirely isolated from its environment. But strictly speaking, no system has exactly autonomous dynamics either: autonomy, like isolation, is always an idealisation, and manifestly there are many systems studied in quantum theory which are satisfactorily idealised as isolated (and unitary).

This is not, however, a serious problem for CH. What matters in their argument, and what does the bulk of the work in their paper, is the physical argument that there are plentiful examples of autonomous quantum systems that are not isolated, and the formal argument that subsystems of unitary quantum systems can display autonomous but non-unitary dynamics. I endorse both claims in the current chapter (indeed, CH's formal argument is closely related to my argument in section 4.2, though CH are less concerned with the mechanisms that ensure autonomy); thus far, the Open Systems View seems fully compatible with the present view. (And CH provide an invaluable explication and defence of some of the metaphysics of open systems, and its interaction with issues like the quantum measurement problem, most which is applicable here also.)

Looking at the details of CH's view, however, reveals very substantial points of disagreement, which can be summarised as saying that CH appear to reject most or all of the four principles of quantum reductionism.

The first principle is that non-unitarity implies interaction. CH are equivocal about this principle: they note (section 5.3) that non-unitary terms in a system's dynamics 'are most naturally interpreted as due to the system's interaction with an external system', but are willing to countenance non-unitary dynamics for the cosmos as a whole and in the context of dynamical-collapse theories. (As an aside: as we have seen, renormalisation means that we should not think of interactions as simply introducing non-unitarity: some part of a system's unitary dynamics might also be explained by its environment.)

CH *seem* to deny the second principle, that interaction is derived from dilation. (And insofar as they do, they effectively deny the third and fourth principles, which presume the second.) They say explicitly (section 2.2) that

> [u]nlike [unitary quantum mechanics], for which the dynamics of an open system of interest, $\mathcal{S}$, is obtained via a contraction … of the total dynamics of $\mathcal{S} + \mathcal{E}$ to the state space of $\mathcal{S}$; in [the Open Systems View], the dynamical equations that govern the evolution of $\mathcal{S}$ pertain directly to $\mathcal{S}$ itself. Systems are modelled as genuinely open … we do not describe the influence of the environment on $\mathcal{S}$ in terms of an interaction between two systems, but instead represent the environment's influence on the dynamical equations that we take to govern the evolution of $\mathcal{S}$.

I quote at length because this seems to encapsulate much of the essence of CH's view, and also because I find it somewhat difficult to interpret. CH seem to be saying that a dynamical equation describing the evolution of $\mathcal{S}$ in the presence of its environment 'pertains directly to $\mathcal{S}$ itself' if it is posited directly, but not if it is derived from contraction of the dynamics of $\mathcal{S} + \mathcal{E}$. But at the formal level, the dynamics of $\mathcal{S}$ pertains to $\mathcal{S}$ simply by virtue of *being* an autonomous dynamics for $\mathcal{S}$: that is, being a rule that determines the future state of $\mathcal{S}$ from its present and past states, perhaps given some broad-brush information (e.g., temperature) about $\mathcal{E}$ but without regard to $\mathcal{E}$'s fine-grained state. And at the physical level, surely what it means for a term in the equation to pertain to $\mathcal{S} + \mathcal{E}$ is that it represents the effect of $\mathcal{E}$ on $\mathcal{S}$: if, physically, some interaction pertains only to $\mathcal{S}$, it's difficult to see how it can represent the effects of something other than $\mathcal{S}$.

More seriously: if we drop the idea that the effects of an environment on a system can be analysed in terms of dynamical couplings between the two, it becomes difficult—practically but also conceptually—to distinguish intrinsic non-unitarity from the effects of the environment, and to understand just how it is that the environment influences the system. To say that a system is influenced by the environment is normally to say something, even if schematically, about how that influence occurs and how it might be manipulated, and to say that quantum theory seems to require us to model the dynamical interaction between the two—which is what CH seem to say we should not do.

In quantum computation, for instance, it is of great importance to minimise the effects of environment on system. Doing so requires us to understand the physics of those effects, so as to manipulate them: we want to know, for instance, whether and how environment-induced non-unitarity depends on environmental temperature, and whether various forms of screening might reduce the environment's effects. Doing so—at least in the standard practice of physics, but arguably also in principle—requires us to model system plus environment together, so that the induced dynamics of the system can be related to the environment's macrostate and the form of the interaction between the two. (It matters, of course, that

the practice of modelling system–environment interactions in this way is highly successful: physics goes far beyond the phenomenological here.)

Similarly, in testing dynamical-collapse theories, we have seen already that the name of the game is to identify and eliminate environment-induced decoherence, so that any residual non-unitarity must be regarded as intrinsic. If the effects of the environment are not to be analysed in the usual way via dynamical interactions, it is opaque—to me at any rate—how even in principle we could obtain evidence for intrinsic non-unitarity.

Finally, there is a question of self-consistency for the Open Systems View thus understood. Suppose $\mathcal{S}$ and $\mathcal{T}$ are both open systems. There is a system $\mathcal{S}+\mathcal{T}$ that comprises both $\mathcal{S}$ and $\mathcal{T}$; it, too, presumably is an open system. The dynamics for $\mathcal{S}$ alone, and $\mathcal{T}$ alone, will be derivable from the joint dynamics of $\mathcal{S}$ and $\mathcal{T}$, and that derivation will reveal the dynamics of $\mathcal{S}$ and $\mathcal{T}$ separately to be, after all, derived from contraction of the joint dynamics of $\mathcal{S}+\mathcal{T}$.

For these reasons, I doubt that it is viable for the Open Systems View to reject the second principle—and, partly because of this, I am unclear as to whether CH intend to do so. If after all we interpret them as *accepting* the second principle, but continue to take them seriously about the inessential role played by unitary dynamics in particular, then we would obtain a coherent version of the Open Systems View which rejects the third principle (that dilations are always to larger, closed systems). In a view like this, we can always dilate a system in order to study its interactions with its environment, but the dilated system in general will itself be an open system—there is no expectation that eventually this process terminates in a closed (i.e., unitarily evolving) system.

A view of this kind could be combined either with the denial or the acceptance of the fourth principle (that the systems to which we dilate are, outside exotic situations, governed by known dynamical laws). If the fourth principle is denied, we end up with an anti-reductionist position that nonetheless has a consistent methodology for scientifically investigating the dynamical causes of non-unitarity in any given system. If it is accepted, we get a view where the full dynamics of any given system is a combination of known and unitary microdynamics plus a non-unitary interaction term; the latter term can be analysed by dilation, but only to a system which in turn has non-unitary interactions, with the whole process continuing without limit.

Both of these views seem coherent. The problem with either is inductive: *in fact* physics, most notably in the (overlapping) realms of statistical mechanics, condensed matter, and effective field theory, has had remarkable success explaining the quantitative dynamics of open quantum systems by dilation to larger *closed* systems, and by assuming that at sufficiently short length scales systems are governed by unitary, and largely known, dynamics. This both provides support for the methodological validity of all four principles, and would be difficult to explain without their metaphysical validity.

At this point I should note that CH give several concrete examples of how rejecting the primacy of unitary dynamics supposedly might benefit physics. For reasons of space I will not discuss all of the examples (the case of quantum reference frames is especially interesting and merits further attention), but I will discuss two of their main examples: the increased generality of the Open Systems equations for non-unitary evolution, and the evaporation of black holes.

The first of these is best understood by looking again at equation (11), the Lindblad equation for non-unitary evolution of a density operator. As I noted, any autonomous Markovian dynamics obtained for a system by tracing over the environment can be put in Lindblad form. But as CH point out, this is not the most general (probability-preserving, Markovian) dynamics that can be written down for a density operator: there exist open-system dynamical equations that cannot be generated from dilation to a larger closed system.

But this possibility is not an advantage for the Open Systems View. Ultimately, it is a matter of experiment which dynamical equation applies to a system. If at some point we were to empirically confirm that there are open quantum systems whose evolution cannot be derived from a larger unitary system, there would be no further need to discuss the respective philosophical merits of the Open Systems View and a more conventional reading of quantum mechanics: that 'more conventional reading' would have been flat-out disconfirmed. But by the same token, the mere logical possibility of that disconfirmation does not speak in favour of the Open Systems View. Indeed, at least methodologically the reverse is the case: precisely because of the profound significance of a disconfirmation of orthodoxy (as much because of its incompatibility with the Standard Model as because of its non-unitarity *per se*), it is advisable to try hard to model a theory within the standard framework and only to abandon doing so if it proves unviable. (This is just the standard principle that one eliminates the possibility of an explanation within extant physics before positing new physics.)

As for the second example—black hole evaporation—Hawking (1975) proposed that the evaporation process was non-unitary; CH observe that this is an entirely legitimate possibility on the Open Systems View and that (section 1) at least part of the motivation for wanting to reject it seems to amount to nothing more than that it runs counter to quantum theory and the fundamental dynamics described by it [on a Closed Systems View]'.

At some level this is a semantic dispute: we could take intrinsic (i.e., not environment-induced) non-unitarity to be included in quantum mechanics, or we could take it as a modification of quantum mechanics. (On either approach, *experimental* evidence for intrinsic non-unitarity would require prior elimination of environmental factors, but that does not apply in this theoretical context.) What matters is whether physicists were disinclined to consider Hawking's proposal due in part to an in-principle rejection of non-unitary dynamics, and it is not at all clear to me that this is the case—of course many, many people have commented on

the information-loss paradox, and some of those comments (especially in popular and semi-popular articles and books) have claimed a contradiction with quantum mechanics, but more detailed and technical arguments seem to have carried more weight—notably Banks, Peskin, and Susskind's 1984 argument that non-unitary effects would be magnified to empirically-ruled-out scales by loop corrections in quantum field theory, and more decisively, the evidence for AdS/CFT duality (Maldacena 1998; Witten 1998; cf., Wallace 2020b and references therein for a conceptual discussion), which if correct shows that black hole evaporation is dual to a manifestly unitary process.

In any case, the choice between quantum reductionism and the Open Systems View is largely empirical, and rests on whether to accept my claim that reductionism (in the weakish sense I intend) is indeed underwritten by the successes of physics. Even if this claim is granted, though, it has one interesting exception, one place where even a reductionist analysis of open quantum systems potentially needs resources not available in unitary quantum mechanics: the problem of the arrow of time in non-equilibrium statistical mechanics.

4.5 The Arrow of Time

Unitary dynamics is reversible: the Schrödinger equation can be run backwards in time as easily as forwards, and there is no intrinsic difference between the two. This is not true for non-unitary dynamics: an autonomous but non-unitary dynamics is irreversible, in the sense that two distinct density operators evolving under that dynamics approach one another in trace norm. Physically (and assuming a dilation-based description of the system–environment interaction) this corresponds to information loss: the system gets entangled with the environment; taking the partial trace discards this information. In any case, the future-direction in time is picked out as the direction of information loss and of convergence in trace norm.[4]

It is then in principle impossible to find an *exceptionless* derivation of autonomous dynamics for $\mathcal{S}$ from unitary dynamics for $\mathcal{S} + \mathcal{E}$, except in the special case where $\mathcal{S}$ is not just autonomous but itself unitary. And so some initial assumption must have been made in the 'derivation' sketched in section 4.2. It is easy to identify that assumption: I assumed that the third term in the Nakajima–Zwanzig equation (3) vanished, something I argued for on the grounds that the microscopic degrees of freedom of the environment 'effectively equilibrated'. Of course, equilibration is itself a time-directed process, and the assumption that a given state equilibrates amounts to the assumption that the state lacks the delicate correlations that would prevent equilibration (see Wallace (2023) for more on this point).

The conclusion of all this is that deriving autonomous dynamics from unitarity requires an assumption about the initial state of system-plus-environment—a harmless-seeming, intuitively reasonable assumption, perhaps, but an assumption which holds for *initial* states but on pain of contradiction cannot hold in time-reversed form for *final* states. In most approaches to the foundations of statistical mechanics, it is this initial-state assumption that underpins irreversibility and which must be explained to explain irreversibility: proposed explanations have included an operationalist (and itself inherently time-directed) requirement that systems must always be thought of as prepared by an agent who is unable to prepare them in sufficiently complicated states to prevent effective equilibration (Myrvold 2021), a (time-directed) principle of causal reasoning that legitimates certain no-correlation assumptions for initial but not final states (Maudlin 2007), and a cosmological boundary condition of one kind or another (Penrose 1994, Albert 2000, Wallace 2023).

There is, however, another tradition in foundations of statistical mechanics: the *interventionist* tradition which seeks to explain irreversibility by the effects of an external environment. Applied to our combined system $\mathcal{S}+\mathcal{E}$, an advocate of this tradition would point out that (i) realistically $\mathcal{S}+\mathcal{E}$ is itself not completely isolated, and (ii) adding even a small amount of non-unitarity to the combined dynamics of $\mathcal{S}+\mathcal{E}$ probably suffices to destroy the delicate correlations that would allow any given initial state *not* to proceed effectively to equilibrium with respect to the microscopic degrees of freedom of $\mathcal{E}$.

The majority view in foundations of statistical mechanics (defended explicitly by, e.g., Sklar 1993: 250–54) is that interventionism is not the right way to explain irreversibility, in part not because the argument is incorrect in itself but because it just pushes the problem of explaining irreversibility back one step, leaving unanswered the *general* question of why systems are so aligned as to produce a consistent direction of irreversibility. I will not attempt to resolve this controversy here. But it is worth observing that environment-based accounts of irreversibility fit very naturally into the Open Systems View—especially into the rather modest version of that view I discussed at the end of section 4.4, which endorsed the first, second, and fourth principles of quantum reductionism while dropping the requirement that autonomous dynamics are always explainable via dilation to a *closed* system. Here at least, the advocate of an Open Systems View has a potential rejoinder to the inductive argument based on the success of closed-system, reductive analyses of autonomous dynamics.

4.6 Conclusions

If quantum mechanics were to be understood first and foremost as an exactly-true theory of the universe, with its applications to subsystems and to restricted

scales understood as derivative or instrumental, then it would be natural to assume that unitarity, and the use of pure rather than mixed states, were fundamental. But there is no reason to do this: the Fundamentality Assumption (that physical theories should be understood as if they were exactly true) and the Cosmological Assumption (that physical theories should be seen as modeling the whole universe) have no justification in scientific practice or naturalised epistemology. Outside a few explicitly cosmological contexts, our goal in physics (and specifically in quantum physics) is in general to model the dynamics of autonomous, though not necessarily isolated, subsystems of the larger universe.

Thus understood, quantum mechanics inevitably leads to the conclusion that quantum states need not be pure and that quantum dynamics need not be unitary. Quantum statistical mechanics provides many examples in which a subsystem of a larger unitary system displays autonomous dynamics not because its degrees of freedom are simply uncoupled to those of the rest of the system but because scale separation and effective equilibration of the system's environment leads to autonomous, but non-unitary, dynamics for the subsystem.

This does not, however, imply a simple pluralism of dynamics in which unitary evolution has no special place. Physics is committed, and justifiably so, to four broadly reductionist principles that can be seen as a naturalised replacement for the unjustified Fundamentality and Cosmological Assumptions, namely (i) that non-unitary dynamics should be explained (except as a last resort) by interaction with an environment; (ii) that 'interaction with an environment' should be mathematically modelled through dilation to a larger quantum system; (iii) that for the purposes of modelling an autonomous subsystem it always suffices to dilate to a *unitary* system; (iv) that we should seek to analyse any such system, ultimately, in terms of the known, and unitary, dynamics of microscopic physics: nonrelativistic particle mechanics, QED, the Standard Model, and so forth. (i) is largely justified methodologically (non-unitarity is evidence for *intrinsic* non-unitarity only once interaction with an environment is ruled out). (ii) is constitutive of what it means for a term in the dynamics of a system to represent interaction with other systems. (iii) and (iv) are not true *a priori*, but receive very strong evidential support from the manifold successes of quantum mechanics, especially in statistical mechanics, condensed-matter physics, and effective field theory.

Cuffaro and Hartmann, in developing their Open Systems View, also reject the Fundamentality and Cosmological Assumptions, and also centre the principle that autonomous quantum systems often evolve non-unitarily. Their view differs from the view here by its rejection of at least some of these four principles of quantum reductionism, though it is to some degree a matter of interpretation just which principles they wish to reject. At least on my reconstruction of their views, it is a semantic matter whether or not they accept the first principle, and a methodological necessity for them to accept the second principle. The status of the fourth principle depends on one's assessment of the success of reductionist methods

(in the moderate sense I intend for 'reductionist') in modern quantum mechanics. The third principle is most interesting: its acceptance or rejection is tied to how the problem of irreversibility in statistical mechanics is to be solved.

Acknowledgements

Thanks to Michael Cuffaro and Stephan Hartmann for the stimulating work that inspired this paper, and for helpful correspondence; thanks also to two anonymous reviewers for thoughtful comments on the draft version of the paper.

Endnotes

1. For evidence that the two assumptions are widespread, see Wallace (2022a: section 1). There are exceptions: notably, the conjunction of the two assumptions gives us something close to Ruetsche's 2011 notion of a 'pristine interpretation' of a quantum theory, and the alternative that she proposes is close to what I propose here. See also Williams (2019), who proposes something similar in the context of effective field theories, though focussed more on rejection of fundamentality than of cosmology.
2. The analysis I provide here is one of two main routes in quantum statistical mechanics to derive autonomous dynamics via scale separation. The other—the quantum version of the so-called BBGKY hierarchy—seeks to develop not an autonomous dynamics of subsystems but an autonomous dynamics of N-point correlation functions for some fixed (usually small) N. See Wallace (2016: sections 2–4, 10) for a conceptual discussion of this latter approach; see also Calzetta and Hu (2008: ch. 1) for a comparison.
3. CH distinguish between (i) the Open Systems View, which is the general position that open systems are prior to closed systems in science, and (ii) the 'general quantum theory of open systems (GT)', which (if I interpret them correctly) they take as an implementation of the Open Systems View in the particular context of quantum mechanics. Since my concern here is exclusively with quantum mechanics, I use 'Open Systems View' to refer to CH's quantum-mechanical position; for ease of comparison with CH's own paper, read 'Open Systems View' as 'GT'.
4. Note that this distinction is independent of whether the unitary dynamics being considered is time-reversal-invariant in the usual sense: even time-reversal-noninvariant unitary processes are reversible.

Bibliography

Albert, D. Z. (2000), *Time and Chance* (Cambridge, MA: Harvard University Press).

Anderson, P. W. (1972), 'More is Different: Broken Symmetry and the Nature of the Hierarchical Structure of Science', in *Science* 177: 393.

Banks, T., Susskind, L., and Peskin, M. E. (1984), 'Difficulties for the Evolution of Pure States into Mixed States', in *Nuclear Physics B* 244: 125–34.

Bassi, A., Dorato, M., and Ulbricht, H. (2023), 'Collapse Models: A Theoretical, Experimental and Philosophical Review', in *Entropy* 25: 645.

Boyanovsky, D. (2015), 'Effective Field Theory out of Equilibrium: Brownian Quantum Fields', in *New Journal of Physics* 17: 063017.

Breuer, H.-P., and Petruccione, F. (2007), *The Theory of Open Quantum Systems* (Oxford: Oxford University Press).

Calzetta, E. A., and Hu, B.-L. B. (2008), *Nonequilibrium Quantum Field Theory* (Cambridge: Cambridge University Press).

Cuffaro, M. E., and Hartmann, S. (2024), 'The Open Systems View', in *Philosophy of Physics* 2: 6.

Hawking, S. (1975), 'Particle Creation by Black Holes', in *Communications in Mathematical Physics* 43: 199.

Maldacena, J. M. (1998), 'The Large N Limit of Superconformal Field Theories and Supergravity', in *Advances in Theoretical and Mathematical Physics* 2: 231–2.

Maudlin, T. (2007), 'What Could Be Objective About Probabilities?', in *Studies in History and Philosophy of Modern Physics* 38: 275–91.

Myrvold, W. C. (2021), *Beyond Chance and Credence* (Oxford: Oxford University Press).

Nakajima, S. (1958), 'On Quantum Theory of Transport Phenomena: Steady Diffusion', in *Progress of Theoretical Physics* 20: 948–59.

Nielsen, M. A., and Chuang, I. L. (2000), *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press).

Penrose, R. (1994), 'On the Second Law of Thermodynamics', in *Journal of Statistical Physics* 77: 217–21.

Quine, W. V. O. (1951), 'Two Dogmas of Empiricism', in *Philosophical Review* 60: 20–43.

Ruetsche, L. (2011), *Interpreting Quantum Theories* (Oxford: Oxford University Press).

Sklar, L. (1993), *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics* (Cambridge: Cambridge University Press).

Stinespring, W. F. (1955), 'Positive Functions on C^*-Algebras', in *Proceedings of the American Mathematical Society* 6: 211.

Wallace, D. (2016), 'Probability and Irreversibility in Modern Statistical Mechanics: Classical and Quantum'. To appear in D. Bedingham, O. Maroney, and C. Timpson (eds.), *Quantum Foundations of Statistical Mechanics* (Oxford University Press, forthcoming). Preprint at https://arxiv.org/abs/2104.11223.

Wallace, D. (2020a), 'Lessons from Realistic Physics for the Metaphysics of Quantum Theory', in *Synthese* 197: 4303–18.

Wallace, D. (2020b), 'Why Information Loss is Paradoxical', in N. Huggett, K. Matsubara, and C. Wüthrich (eds.), *Beyond Spacetime: The Foundations of Quantum Gravity* (Cambridge: Cambridge University Press), 209–36.

Wallace, D. (2022a), 'Isolated Systems and their Symmetries, Part I: General Framework and Particle-Mechanics Examples', in *Studies in the History and Philosophy of Science* 92: 239–48.

Wallace, D. (2022b), 'Quantum Gravity at Low Energies', in *Studies in History and Philosophy of Science* 94: 31–46.

Wallace, D. (2023), 'The Logic of the Past Hypothesis', in B. Loewer, B. Weslake, and E. Winsberg (eds.), *The Probability Map of the Universe: Essays on David Albert's* Time and Chance (Cambridge, MA: Harvard University Press), 76–109. Originally published online in 2011 (http://philsci-archive.pitt.edu/8894/).

Wallace, D. (2025), 'Thermodynamics With and Without Irreversibility'. In O. Lombardi and C. Lopez (eds.), The Arrow of Time: From Local Systems to the Whole Universe (Cambridge University Press, 2025), pp. 12–55.

Williams, P. (2019), 'Scientific Realism Made Effective', in *British Journal for the Philosophy of Science* 70: 209–37.

Wilson, M. (2006), *Wandering Significance: An Essay on Conceptual Behavior* (Oxford: Oxford University Press).

Witten, E. (1998), 'Anti de Sitter Space and Holography', in *Advances in Theoretical and Mathematical Physics* 2: 253–91.

Zwanzig, R. (1961), 'Memory Effects in Irreversible Thermodynamics', in *Physical Review* 124: 983.

Zwanzig, R. (1965), 'Time-Correlation Functions and Transport Coefficients in Statistical Mechanics', in *Annual Review of Physical Chemistry* 16: 67–102.

5

Frameworks in Physics: Abstractness, Generality, and the Role of Metaphysics

Doreen Fraser and Adam Koberinski

5.1 Introduction

In physics, there are models and theories, but also frameworks. The explicit focus of Cuffaro and Hartmann (2024) is a set of arguments that aim to show that open systems are more fundamental than closed systems in quantum theory, but in the course of making these arguments they articulate a conception of frameworks as operating at a more general level than theories (and models). While the idea that frameworks play a role in physics is not new, we regard Cuffaro and Hartmann's account as a valuable original contribution to the recent literature on this topic. In our contribution to this volume we will critically assess Cuffaro and Hartmann's account of frameworks and analyse two other examples of frameworks in contemporary physics: (relativistic) quantum field theory (QFT) and (non-relativistic) statistical mechanics. Our main disagreement with Cuffaro and Hartmann lies in our argument that it is not appropriate to assess metaphysical claims about fundamentality at the framework level, but only at the theory or model levels. While our conclusions have implications for their position on the fundamentality of open systems, evaluating this thesis is not our main goal. In the final section we will briefly discuss the status of open and closed systems in the applications of the frameworks that we consider to quantum systems.

Over the past several decades, the models literature within philosophy of science has focused attention on the important roles that models play in science. This has involved distinguishing models and their functions from theories. For example, in their own contribution to their influential volume *Models as Mediators*, Morrison and Morgan (1999) emphasise that the autonomy of models is essential for performing their function of mediating between theory and world. Theories and data are used in the construction of models, but models are partially independent of both. Roughly speaking, frameworks stand to theories as theories stand to models. A survey of discussions of frameworks in the models literature is beyond the scope of this chapter, but for our purposes it is interesting to note that Hartmann (1998) offers early reflections on the framework–theory–model distinction that

Doreen Fraser and Adam Koberinski, *Frameworks in Physics: Abstractness, Generality, and the Role of Metaphysics*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press. © Oxford University Press (2026). DOI: 10.1093/9780198929277.003.0006

are informed by an analysis of QFT. In the course of arguing that idealised models serve the cognitive function of facilitating the exploration of features of a theory, Hartmann characterises Type A Theories as general background theories and cites the general formalism of QFT (e.g., as characterised by the Wightman axioms) as an example, which is in agreement with our own construal and examples of frameworks below. Type A Theories (or frameworks) are distinguished from Type B Theories that adequately describe some relatively broad-scope but limited domain such as QED and QCD as well as from Type A Models that are toy models such as the ϕ^4 model, which are all examples of theories on our account. Type B Models are phenomenological models. From a contemporary perspective, this set of distinctions was prescient. Subsequent philosophical investigations of QFT have reinforced the moral that the framework of QFT is distant from both the particle physics theories (e.g., quantum electrodynamics (QED), quantum chromodynamics (QCD), the Standard Model) and models (e.g., of specified target systems). As we will articulate below, frameworks are more general and abstract than the theories they subsume, which in turn are more general and abstract than the models that connect theory to phenomenology. In the mainstream approach to QFT, the variety of renormalisation and regularisation techniques invoked to construct models with varying degrees of success in different contexts have underscored that it is a nontrivial achievement to go from a framework to a model (via theories). Similarly, within the axiomatic approach to QFT, the constructive QFT programme of constructing models of some set of axioms for specified interactions has proven to be a challenging project in mathematical physics.

Cuffaro and Hartmann roughly distinguish models, theories, and frameworks as 'ordered according to their generality, i.e., with regard to how "far away" each is, conceptually, from the so-called target system in a given field of inquiry' (Cuffaro and Hartmann 2004: 5). Differentiation by levels of generality is a widely-agreed-upon starting point for philosophical analysis of frameworks, theories, and models, but it leaves many details to be specified. Cuffaro and Hartmann fill in the details by supplying a careful analysis of two examples of frameworks—standard quantum theory (ST) and general quantum theory of open systems (GT)—that highlights what they take to be the key features of frameworks. These frameworks are regarded as alternative formulations of quantum theory. We will follow Cuffaro and Hartmann in using examples to explicate the framework–theory–model distinction, but we will focus on different examples. These examples illustrate additional important features of frameworks that are not part of Cuffaro and Hartmann's explicit analysis of frameworks.

The two frameworks that we will examine are (relativistic) QFT and (non-relativistic) statistical mechanics.[1] Our purpose in introducing different examples of frameworks than Cuffaro and Hartmann's is to shed light on other features and functions of frameworks in contemporary physics. Unlike the ST and GT frameworks for quantum theory, the QFT and statistical mechanics frameworks

are not alternative formulations of one framework. Both the QFT and statistical mechanics frameworks can be used to construct quantum models of some systems. However, we do not contend that we have uniquely identified the correct frameworks for quantum physics. We will follow Carnap's counsel of tolerance towards linguistic frameworks in our approach towards frameworks that are not solely linguistic: '[l]et us be cautious in making assertions and critical in examining them, but tolerant in permitting linguistic forms' (Carnap 1983: 257). More specifically, since the theories that fall under frameworks overlap in nontrivial ways, there are different natural choices for carving up physics into frameworks. To give another example, Wallace (2018, 2020) takes quantum theory as a whole to be one framework, within which QFT and non-relativistic QM are sub-frameworks. The contrasting framework in that case is classical theory. But one might take Lagrangian theories to constitute a framework instead, under which both classical and quantum theories can be formulated. Of course, Carnap's recommendation of tolerance is limited: some proposed frameworks may prove not to be useful and get abandoned.[2]

Our focus will be on critically examining Cuffaro and Hartmann's account of 'views', which include metaphysical assumptions about the nature of the objects in the domain, as being integral to the formulation of frameworks (Cuffaro and Hartmann 2004: 5). On our account, frameworks are characterised by their degree of abstractness—in the sense of the opposite of concreteness—as well as their generality of scope—in the sense of applying to a wide range of phenomena. Our thesis is that analysis of the QFT and statistical mechanical frameworks as well as the ST and GT frameworks indicates that metaphysical assumptions are not integral to the content of frameworks; instead, metaphysical commitments may enter at the theory and model levels of representation. One of the pragmatic virtues of frameworks is their abstractness; many different metaphysical commitments may be added in the step of moving from thin physical interpretation to a more concrete theory or model. In Section 5.2 we describe the statistical mechanics and QFT frameworks. In Section 5.3 we compare and contrast our account of the statistical mechanics and QFT frameworks to Cuffaro and Hartmann's account of the ST and GT frameworks. In Section 5.4 we briefly consider how the QFT and statistical mechanics frameworks are used to represent open and closed quantum systems.

5.2 Examples of Frameworks: Non-Relativistic Statistical Mechanics and Relativistic QFT

One broad motivation for carving up frameworks in contemporary physics along non-relativistic vs. relativistic lines is that the division corresponds to one of the major revolutions in twentieth-century physics. As noted, an alternative carving

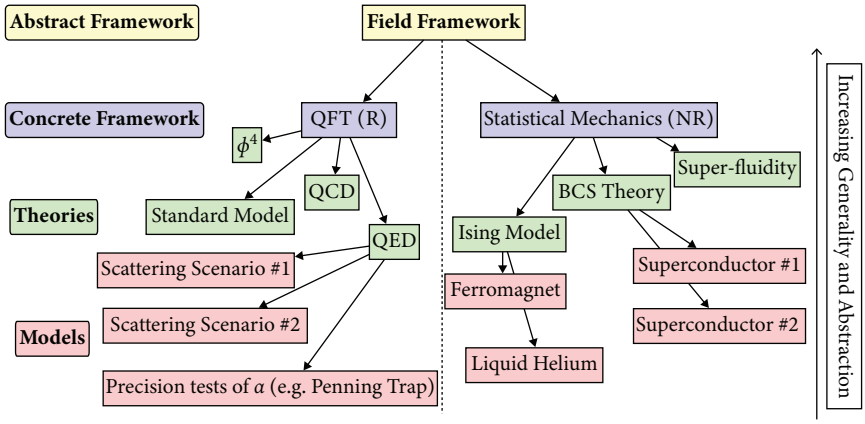


Figure 5.1 Some examples of concrete frameworks, theories, and models falling under the abstract field framework. Of particular focus here is the division between (relativistic) QFT and (non-relativistic) statistical mechanics as frameworks incorporating different physical interpretations (e.g., relativistic vs. non-relativistic spacetime structure). The frameworks sit at higher levels of generality and abstraction than theories, which in turn are more general and (typically) abstract than models.

slices along the lines of quantum vs. classical physics, the other major revolution. A more specific historical motivation for examining non-relativistic statistical mechanics and relativistic QFT as parallel frameworks is that analogies between models constructed using these frameworks have led to important discoveries. Our account in this section is informed by the case studies in Fraser and Koberinski (2016) and Fraser (2020), and the account of QFT in Koberinski (2019). We will begin by explicating these examples in order to set up our general philosophical characterisation of frameworks at the end of this section.

Both the QFT framework and the statistical mechanics framework fall under the more abstract *field framework* (see Fig. 5.1). While it is not strictly necessary to identify the field framework level, it is helpful to recognise that the QFT and statistical mechanical frameworks have some common elements. For example, they share the representational strategy of representing a large number of degrees of freedom using mathematical fields $\phi_i(x)$ over some stratum x. $\phi_i(x)$ is not necessarily scalar-valued; it may be vector-valued or operator-valued. x may be either continuous or discrete. Field values fluctuate; physically significant quantities take the form of products of fields evaluated in privileged states (e.g., vacuum expectation values or correlation functions). Another ingredient of the field framework is the use of a Lagrangian or Hamiltonian to represent the dynamics. Renormalisation group (RG) methods are an example of a representational and calculational component (Koberinski and Fraser 2023).

The framework of QFT is a field framework that is mainly used for high-energy particle physics. Some of its core conceptual features may include[3] fields defined on Minkowski spacetime, a privileged vacuum state, vacuum expectation values, local interactions, cluster decomposition, functional integrals, and perturbative renormalisation techniques. These are a mix of physically motivated principles that underlie theories formulated within the framework, methodological strategies for formulating theories and constructing models, and mathematical tools for derivation and calculation.

The statistical mechanics framework is a field framework that is typically applied to represent condensed matter systems. A significant difference from the QFT framework is that the mathematical and methodological strategies are applied to describe non-relativistic causal structure. The space(time) structure is usually taken to be either Galilean or Euclidean. There is a global time variable, simultaneity is absolute, and there is no upper limit on the speed of signal propagation. This crucial difference has several downstream implications for the implementations of each framework, both conceptual and calculational. For example, there is an absolute standard of simultaneity that makes the use of spatial representations, such as an array of spins on a spatial lattice of atoms to model a ferromagnet, straightforward. Importantly, statistical mechanics can be used to construct either quantum or classical theories of field systems. When the field variables are defined over a continuum they are often regarded as an idealisation, since many systems treated statistically in this way are taken to be composed of many discrete degrees of freedom (e.g., a lattice of atoms in a ferromagnet). Philosophical discussion of the statistical mechanics framework has focused on coarse-graining, which involves systematically transforming from a theory that includes all fine-grained details relevant at one scale to different theories which exclude fine-grained details that are negligible at larger distance scales. Coarse-graining is a physical interpretation of the application of renormalisation group methods which is specific to the statistical mechanics framework; this aspect of the concrete physical interpretation of the statistical mechanics framework is not shared with the QFT framework (see Koberinski and Fraser 2023 for elaboration and analysis).

Each framework can be used to formulate a variety of theories that each describe large classes of phenomena. In the QFT framework, examples of theories include QED, QCD, the electroweak theory, and ϕ^4 theory. In statistical mechanics, examples of theories include the Ising model for ferromagnetism, theories of superconductivity (like the BCS theory), and superfluidity. These are closest to what philosophers have traditionally called theories. Models for each theory are typically not strictly derived from the theory, but have some phenomenological input to connect the more abstract theory to a specific set of phenomena.

In common with Cuffaro and Hartmann's characterisation of the tripartite framework–theory–model distinction, these frameworks are more general than

theories, which are in turn more general than models, in the sense of scope of applicability. A framework is more general than a theory in that a larger class of physical phenomena can fit under the framework than the theory. Similarly, theories are more general than models in this sense. Furthermore, the levels in Fig. 5.1 represent increasing levels of *abstractness* from models to frameworks. A decrease in the level of abstractness means that concrete physical interpretation is added. For example, the statistical mechanics framework introduces a concrete physical interpretation of the fields as representing degrees of freedom arranged in non-relativistic space(time). Moving down one more level involves further interpreting the degrees of freedom as being atoms, electrons, photons, or phonons, for example. On the other side of Fig. 5.1, moving from the abstract field framework level to the concrete QFT framework involves specifying that the fields used to represent degrees of freedom are operator-valued, that vacuum states are privileged, and that both figure in vacuum expectation values. Moving down another level, in QED expressions involving vacuum expectation values can be given an empirical interpretation in terms of scattering amplitudes. In QCD, quark confinement adds to the empirical and physical interpretation in a different way.

Generality and abstractness are related, but conceptually distinct. Generality is a measure of the scope of applicability of a framework, theory, or model: the more general it is, the more phenomena fall under it. Abstractness, on the other hand, is a measure of the degree to which physical interpretation is specified. In physics, it is often the case that a greater degree of generality is achieved by making something more abstract, but the two are at least conceptually distinct. Increasing abstraction as one gets to the framework level results in greater generality, but also less physically contentful descriptions. For the frameworks discussed here, there are high-level physical principles, spacetime structure, and the mathematical machinery of fields that allow one to treat large numbers of degrees of freedom. These are all physically motivated, but at the level of the framework are so abstract as to admit a range of possible concrete physical interpretations. Even the field framework common to both contains abstract physical content (e.g., the concept of a degree of freedom); it is still applied mathematics, not pure mathematics. However, the physical descriptions afforded by frameworks are usually so thin that most questions about the ontology (of a represented domain) are only answerable at the more concrete levels of theories or models, when a representational modelling strategy is specified. Note that abstractness as opposed to concreteness aligns with traditional definitions of the term, but differs from other definitions offered in the context of the recent philosophical debate regarding explanation in paradigm applications of the statistical mechanics framework such as critical phenomena (e.g., Morrison 2015: 20; Reutlinger and Andersen 2016; Jansson and Saatsi 2019; and Batterman 2019). This debate has centred on the role of coarse-graining in applications of the statistical mechanics framework, which as noted above is an aspect of the concrete physical interpretation

of the statistical mechanics framework. The QFT framework is not given a concrete physical interpretation in terms of coarse-graining, and the field framework is more abstract (in our sense) because it refrains from adding a physical interpretation of the coarse-graining type or any other type (see Koberinski and Fraser 2023 for details).

One of the benefits of a good framework is an optimal trade-off between specificity and the flexibility it allows for modelling diverse types of physical systems. The specificity is needed to provide direction for choosing an appropriate formalism and a template for modelling, while the flexibility is needed to allow a larger class of systems to fall under the framework. The payoff of a good framework is that it facilitates the transfer of successful modelling strategies from one domain to another. For example, RG methods are a family of techniques used to derive physically and empirically significant quantities in models that apply the field framework. One indication that RG methods might be a suitable problem-solving strategy is that field values fluctuate over a wide range of scales. Scale is an abstract physical concept that can be given different interpretations (e.g., space, energy, time) in different models. While in each specific case the scale takes a specific form, the abstractness helps one to identify contexts in which some form of the RG can be used. The framework serves as a toolkit for modelling situations in which field values fluctuate and are weakly coupled across that scale. This is a pragmatic virtue that would disappear if the framework is too specific. There is a trade-off between specifying the physical content of the representation so that potential applications for the problem-solving strategy can be recognised and preserving the abstractness that allows a wide range of physical problems to be solved using the same strategy.

5.3 Abstractness, Metaphysics, and Fundamentality

Now that we have explicated our own examples of frameworks, we turn to comparing our account of frameworks with Cuffaro and Hartmann's account. Cuffaro and Hartmann (2024) distinguish *framework*, *theory*, and *model* as 'ordered according to their generality, i.e., with regard to how "far away" each is, conceptually, from the so-called target system in a given field of inquiry' (Cuffaro and Hartmann 2004: 5). This suggests a combination of abstractness and generality as we have defined them. However, the abstractness of frameworks plays a more important constitutive role in our account of frameworks. What is important in Cuffaro and Hartmann's account is that frameworks are formulated with reference to a *view*, which is associated with '(i) a set of methodological presuppositions in accordance with which we characterize the objects of a given domain, that is (ii) motivated by a particular metaphysical position concerning the nature of those objects' (Cuffaro and Hartmann 2024: 5). We agree that frameworks incorporate

methodological presuppositions, but disagree with the central role attributed to metaphysical assumptions. In particular, Cuffaro and Hartmann argue that the closed systems view is the metaphysical position that motivates ST, while the open systems view motivates GT in the sense that 'the subject matter of a given scientific domain is thought of (a priori) to consist of systems that are in general' closed or open, respectively (Cuffaro and Hartmann 2004: 6). While a metaphysical position may motivate the formulation of a framework, we will argue that this metaphysical position need not be retained in the output when the framework is used to construct a model. This is apparent for QFT and statistical mechanics as well as the ST and GT frameworks. We take a consequence to be that frameworks are not the appropriate level at which to assess fundamentality claims.

In the same way that the open and closed systems views are taken to motivate the GT and ST frameworks, respectively, the metaphysical position that the objects of study are fields is the obvious candidate for the view that motivates the field frameworks of QFT and statistical mechanics. Is there an implied commitment to a field ontology in the use of these frameworks to represent systems? The applicability of the framework entails that a system can be *approximated* using systems of mathematical fields, but is there necessarily a further commitment to a field ontology? The answer is no for both the QFT and statistical mechanics frameworks. Consider statistical mechanics. In prototypical examples such as the representation of continuous phase transitions in fluids or ferromagnets, models represent a target system with a large number of discrete degrees of freedom. Mathematical fields $\phi(x)$ are used to represent these degrees of freedom, but physical fields with continuously many degrees of freedom are typically only introduced as an idealisation. The use of mathematical fields as a modelling tool does not entail an ontological commitment to the system consisting of physical fields. While a commitment to fields as modelling tools may be constitutive of the framework, it is not an ontological commitment, and certainly not a commitment to the fundamentality of fields. Idealisations are a widely used modelling strategy for representing a target system. Idealisation allows frameworks to be flexibly applied to a broader range of theories and target systems. In the continuous phase transition case, a payoff of employing the idealisations that allow the field framework to be applied is arguably explanatory power (Batterman 2002).

What about QFT? Here the answer is not so straightforward, but it remains the case that the framework does not determine the answer. The appropriate ontology for QFTs is still up for debate, but there is general consensus that particles are not part of the fundamental ontology (see Fraser 2022 for a review). This leaves many options open, including embracing particles as emergent ontology (e.g., Wallace 2001). One might think that some form of field is the best candidate for the fundamental ontology, but there are also arguments against some varieties of field ontologies (e.g., Baker 2009). What is important for present purposes is that these debates are not settled by noting that fields play a primary role in the *formalism*

of QFT. We cannot read off the ontology of QFT from examining the formalism at the framework level. Further arguments, physical and philosophical, are needed to draw conclusions about ontology. We take this to be a general point of caution in the metaphysical analysis of physics: ontological commitments cannot simply be read off mathematical formalism, especially when that formalism is presented at an abstract level in the context of frameworks. Further specification is needed to determine *what* is being represented, and *how* it is being represented. This is made concrete through physical interpretation in models which represent systems of the same type.

Cuffaro and Hartmann's own analysis in this volume reaches the similar conclusion that the metaphysical commitment associated with a view is not necessarily retained in the application of a framework. They conclude that 'from the point of view of each of these two [orthodox and (neo-)Everettian] families of approaches to interpretation, … although ST is formulated in accordance with the closed systems view, it is nevertheless ontologically committed to open systems' (Cuffaro & Hartmann, this volume, sec. 1). The metaphysical input to a framework is dissociated from its metaphysical output when it gets applied. Moreover, determining the ontology requires going beyond the framework itself by specifying an interpretation of some sort. This is the feature of abstractness that we have made explicit and emphasised in our account of frameworks; here it enters implicitly in Cuffaro and Hartmann's account of the use of the ST and GT frameworks.

The abstractness of frameworks carries further consequences for Cuffaro and Hartmann's arguments for the fundamentality of the open systems view. Cuffaro and Hartmann argue that GT *as a framework* is more fundamental than ST. We contend that the framework level is not the appropriate level at which to assess claims of fundamentality. If relative fundamentality claims are taken to be metaphysical claims, then frameworks are not the appropriate targets for these claims. On one standard understanding of fundamentality, the metaphysical relation of fundamentality is primarily applied to things in the world, though it finds a natural extension to theories and models when we take these latter to represent things in the world. But, on our account, frameworks are too abstract to support the evaluation of relative fundamentality claims. For example, consider the statistical mechanics framework. The Galilean or Euclidean space(time) structure of statistical mechanics, for example, is a metaphysical assumption that is an input to the framework. But there is no case to be made from looking at the framework alone that this space(time) structure is fundamental, or that all systems modelled within this framework are fundamentally Galilean. Again, the use of idealisations or approximations in the construction of a model factors into judgements about the relative fundamentality of the representation. The frameworks provide a formalism and set of concepts within which to construct models, and we learn about the ontology of the world through studying the relationship between models and the world, not the framework and the world.

Cuffaro and Hartmann (2024) implicitly recognise the abstractness of frameworks in their definitions of fundamentality. As a starting point for defining the ontic fundamentality of a framework, they introduce the concept of a 'capital-O' Object of a framework that is common to every theory and model constructed using the framework (Cuffaro and Hartmann 2004: 15). In contrast, a 'little-o' object is a (constituent of) a given system, which is the referent of a model (Cuffaro and Hartmann 2004: 15). Cuffaro and Hartmann draw attention to the abstractness of capital-O Objects, at one point describing the capital-O Objects in the ST and GT frameworks as the abstract concepts of a state vector $|\psi\rangle$ and a density operator ρ, respectively. The proposed definition of ontic fundamentality of frameworks in terms of capital-O Objects is rejected by Cuffaro and Hartmann on the grounds that it is 'too abstract': 'we have not considered the little-o objects—the ontologies—that the capital-O Objects of ST and GT actually represent' (Cuffaro and Hartmann 2004: 17). Again, as Cuffaro and Hartmann's analysis of ST illustrates, the problem is that it is not possible to determine the relevant ontological properties of the 'capital-O' Objects because different interpretations and assumptions that enter in at the model level affect the ontological properties of the 'little-o' objects (e.g., orthodox vs. Everett interpretations). The views associated with a framework do not settle questions about relative ontic fundamentality. Cuffaro and Hartmann offer a revised definition of ontic fundamentality (Ont-Fund-2) that still applies at the framework level (21). This definition breaks the connection between the view associated with a framework and the determination of relative fundamentality and introduces the representation of little-o objects as a determining factor. From our perspective, continuing to apply ontic fundamentality to frameworks is not the right way to proceed. Instead, abstractness needs to be recognised as a central feature of frameworks. As a result, questions about ontic fundamentality are not appropriately posed at the framework level because they cannot be settled at this level.

In some instances it may make sense to talk of the relative fundamentality of different frameworks as a convenient shorthand. This is the case when all of the theories and models constructed using one framework are relatively more fundamental than the theories and models constructed using the second framework. In these cases the conceptual foundations of the less fundamental framework can often be found as limits or approximations to the conceptual foundations of the more fundamental. So, for example, we can say that the framework of quantum theory is more fundamental than classical theory, or that the relativistic framework is more fundamental than the non-relativistic framework. But these are special cases. We can't similarly say that Lagrangian mechanics is more or less fundamental than QFT, nor that the relativistic framework is more fundamental than quantum theory. For QFT and statistical mechanics, only loose relative fundamentality claims can be made, owing to the relativistic structure of QFT not present in statistical mechanics.[4] But the same does not hold for the case of ST

and GT, since they are purportedly rival frameworks covering the same domain of quantum theory.

To summarise, on our account of frameworks their defining features are generality and abstractness. The degree of abstractness of frameworks precludes them from fixing the ontology of the systems which are represented with models constructed using them. For example, the field frameworks of QFT and statistical mechanics do not entail a commitment to field ontology; similarly, the ST and GT frameworks do not automatically imply the fundamentality of closed and open systems, respectively. On our account, this is not an accidental feature of the ST and GT frameworks for quantum theory, but an essential feature of frameworks in general. Frameworks are powerful due to their abstractness and generality, and both of these features depend on assigning minimal physical interpretation at the framework level. Cuffaro and Hartmann argue that the open systems view is more fundamental by arguing that the framework associated with this view, GT, is more fundamental. On our account, it is not appropriate to assert relative fundamentality claims about frameworks. The utility of frameworks lies in their ability to accommodate many different types of systems by getting 'filled-in' with different physical and metaphysical content in theories and models. As a result, theories or models are the appropriate objects of claims about the relative fundamentality of the represented systems (or their constituents). We note that a similar conclusion is drawn by Wallace (2020) in the context of interpreting quantum theory; he takes quantum theory to be a framework, and argues that ontology must be specified relative to particular theories or models constructed within this framework.[5]

These criticisms of the argumentative strategy in Cuffaro and Hartmann (2024) do not bear on the truth of their conclusion that open systems are more fundamental than closed ones. We will consider how the models constructed using the QFT and statistical mechanics frameworks bear on this question in the next section.

5.4 Open Systems in the QFT and Statistical Mechanics Frameworks

Instead of arguing that GT is a more fundamental framework than ST, Cuffaro and Hartmann could instead skip the arguments regarding the frameworks and argue directly that the open systems view is better suited to the actual modelling practices of *both* GT and ST, or to quantum theory in general. In fact, they do provide such arguments (see their chapter in this volume); we take these as being an appropriate way to argue for the fundamentality of the open systems view in quantum theory. Even in frameworks where the paradigmatic modelling assumptions treat a system as closed, it can still be the case that open systems are themselves more fundamental. The positive point we take from Cuffaro and Hartmann is a methodological point about modelling practices: that proper consideration of

frontier physics practice should motivate a methodological switch to treating systems as open by default, with closed systems being a special case. They focus on frameworks of quantum theory, but can the field frameworks discussed here accommodate such a switch? In both cases the answer is a clear yes, with some qualification.

First, we need to clarify what it means for a system to be open. There seem to be three related, though distinct, conceptions at play. First is the idea that a system modelled as open is treated in a context where the environment with which it interacts is also modelled. Second, and somewhat stronger, an open system representation might be one that includes exogenous control variables, implicitly modelling the system as interacting with some part of the environment. Third, and strongest, is a representation that treats interactions with the environment on a par with dynamics internal to the system. We take a permissive understanding here, taking all of these to count as open systems. Due to the flexibility of modelling strategies within and across frameworks, different conceptions may be needed in different contexts.

Start with QFT. Paradigmatic uses of the QFT framework involve the scattering matrix: modelling transition amplitudes between states of free particles at asymptotically early and late times. This typically involves adopting a closed-systems perspective in which one can model the in- and out-states as effectively free particles. But this is only one (admittedly paradigmatic) use of the QFT framework; there are others. In particular, the effective field theory framework involves an explicit split of system and environment in the form of low- and high-energy degrees of freedom, respectively. While the standard uses of effective field theory make strong assumptions about the high-energy degrees of freedom (the environment) to ensure a minimal impact on the system, the split is essential to the definition of an effective field theory. An effective field theory is inherently an open system, with impacts from the environment modelled systematically as higher-dimensional interaction terms of the low-energy fields. While the QFT framework as a whole does not require an effective field theory perspective, this is one major application in which the QFT framework admits of the widespread use of open systems, albeit in the weakest sense of 'open'. The QFT framework is flexible in accommodating different modelling practices within it.

More obviously, finite-temperature field theory (aka thermal field theory) offers a relativistic treatment of open, thermally interacting systems. The general strategy involves combining elements of the QFT and statistical mechanics frameworks. Instead of dealing with isolated excitations to a vacuum state, finite-temperature field theory takes the background state to be a thermal state. For example, the thermal state can be a KMS state defined by its satisfaction of the KMS condition in (non-relativistic) quantum statistical mechanics, which is an infinite-dimensional generalisation of the Gibbs state for a system at inverse temperature β (Haag et al. 1967).[6] One can deal with quantum fluctuations about an equilibrium state, or

more generally with larger departures from equilibrium. Temperature is now an exogenous control variable in finite-temperature field theory, and the degrees of freedom that are the focus of the modelling context are coupled to the background thermal state. Thus finite-temperature field theory uses a hybrid of the QFT and statistical mechanics frameworks to treat systems that are in some sense irreducibly open.

Finally, there is recent work on treating open systems within the effective field theory approach to QFT. In this case, one derives an effective density matrix for the effective field theory, and shows that its evolution satisfies a modified Lindblad equation (Braaten et al. 2016). Effective open QFT combines the openness of standard effective field theory with the dynamical evolution of GT, thereby providing the strongest example of open systems in QFT. While their use remains limited due to the technical complexity of the general calculations, it is clear that the QFT framework admits of many different formalisms relating to open systems.

For statistical mechanics, the systems under study are typically systems for which thermodynamic treatments prove informative of their properties. Statistical mechanical and thermodynamic treatments of systems often deal with open systems, where the system under study is allowed to exchange certain macroscopic quantities (e.g., heat, particles, energy) with an environment modelled as something like a heat bath. While closed systems are also part of statistical mechanics, they are in general a special case. The open systems view is therefore a good fit for the disciplines that use the statistical mechanics framework. Most systems modelled in this framework are taken to be open by default, in line with the methodological conclusion from Cuffaro and Hartmann.

5.5 Conclusion

We have argued that the distinction between frameworks, theories, and models in physics is characterised by both *abstractness* and *generality*. Frameworks are more abstract and general than theories, which are in turn more abstract and general than models. Generality is a measure of scope: a representation is more general when it applies to a larger domain of phenomena. Abstractness, by contrast, is a measure of concrete physical interpretation: a representation is more abstract when its physical interpretation is less elaborated. Both of these qualities are important for the flexibility and usefulness of frameworks: abstractness is often needed to increase generality.

While we agree with some aspects of the characterisation of frameworks laid out in Cuffaro and Hartmann (2024), we argue that the abstract nature of frameworks makes them poor hosts for metaphysical views. Part of the value of a framework lies in its flexibility for modelling many different types of systems; a framework wedded to a particular metaphysical view about the fundamentality of a given

representation would be unnecessarily restricted in its application. Therefore, we argued that the framework level is not the appropriate level at which to argue for the fundamentality of an open systems view.

In our analysis of the field frameworks of QFT and statistical mechanics, we highlighted their generality and abstractness, and their resultant flexibility. We argued that Cuffaro and Hartmann's analysis of the ST and GT frameworks also recognises their generality and abstractness, at least implicitly. As a result, metaphysical and interpretative commitments are applied downstream of the framework, at the level of theories or models. Illustrations of this point are that different kinds of systems—not only fields—can be modelled in the QFT and statistical mechanics frameworks, and that open systems models can be constructed using both frameworks.

Our conclusions are compatible with Cuffaro and Hartmann's thesis that open systems are more fundamental than closed systems in the methodological sense that proper consideration of frontier physics practice should motivate a methodological switch to treating systems as open by default, with closed systems being a special case. The appropriate level at which to argue for relative fundamentality, however, is at a more concrete level than the framework level, where the physical interpretation is sufficiently specified to support ontological claims.

Acknowledgements

We are grateful to Hazel Gifford for assistance with the figure, to two anonymous reviewers for their feedback on an earlier draft, and to Michael E. Cuffaro and Stephan Hartmann for inviting us to contribute this chapter. DF thanks David Wallace for helpful discussions and the audience at the 2022 Grünbaum Memorial Lecture at the Center for Philosophy of Science, University of Pittsburgh. DF and AK gratefully acknowledge support from a Social Sciences and Humanities Research Council of Canada Insight Grant.

Endnotes

1. For the sake of brevity, in this chapter the term 'QFT' will be taken as synonymous with 'relativistic QFT' and to have its paradigmatic applications in particle physics. As we explain in the next section, the field framework is also applied to construct non-relativistic quantum representations of condensed matter systems, but we will refrain from using the term 'QFT' to describe these. For clarity, we use the term 'finite-temperature field theory' to refer to the application of the field framework to construct relativistic representations of thermal systems.
2. Proposed frameworks and their relationships can also be subject to philosophical/foundational analysis. For example, from the perspective of algebraic QFT, it is natural to consider the general framework for quantum theory to be quantum theory with an infinite number of degrees of freedom, which admits unitarily inequivalent Hilbert space representations (see Ruetsche 2011). On

this way of carving things up, Cuffaro and Hartmann's ST framework has limited applicability (e.g., to non-relativistic QM); ST applied to non-relativistic QM and relativistic QFT are related as different applications of the general quantum-theory-with-an-infinite-number-of-degrees-of-freedom framework. In contrast, Cuffaro and Hartmann present ST as a general framework for quantum theory that has both non-relativistic and relativistic applications.

3. We set aside debates about how to characterise QFT both physically and mathematically that are not relevant to our discussion here. Examples of other elements of the framework are the use of Fock space to represent particles and the template for spontaneous symmetry breaking.
4. One can further argue that QFT is in some sense more fundamental than statistical mechanics, but this is only by appeal to the systems that each framework is used to represent (fundamental particles versus atomic-scale matter) or the main theories fitting into each framework (the Standard Model of particle physics versus various condensed matter theories). This serves to illustrate the point that the notion of fundamentality is better suited to less abstract representational content, at the level of theories, models, or the entities themselves.
5. For an opposing view, see Guo (2023).
6. While the original formulation of the KMS condition was in the context of (non-relativistic) quantum statistical mechanics, there has more recently been a generalisation to real-time, relativistic contexts (Birke and Fröhlich 2002). The construction involves analytically continuing 'imaginary time' Green's functions that satisfy the Euclidean KMS condition into 'real time' Green's functions. The analogous condition satisfied by these real-time Green's functions is a relativistic version of the KMS condition.

Bibliography

Baker, D. J. (2009), 'Against Field Interpretations of Quantum Field Theory', in *The British Journal for the Philosophy of Science* 60: 585–609.

Batterman, R. W. (2002), *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence* (New York: Oxford University Press).

Batterman, R. W. (2019), 'Universality and RG Explanations', in *Perspectives on Science* 27: 26–47.

Birke, L., and Fröhlich, J. (2002), 'KMS, etc.', in *Reviews in Mathematical Physics* 14: 829–71.

Braaten, E., Hammer, H.-W., and Lepage, G. P. (2016), 'Open Effective Field Theories from Deeply Inelastic Reactions', in *Physical Review D* 94: 056006.

Carnap, R. (1983), 'Empiricism, Semantics, and Ontology', in P. Benacerraf, and H. Putnam (eds.), *Philosophy of Mathematics: Selected Readings* (Cambridge: Cambridge University Press), 241–57. Second edition.

Cuffaro, M. E., and Hartmann, S. (2024), 'The Open Systems View', in *Philosophy of Physics* 2: 6.

Fraser, D. (2020), 'The Development of Renormalization Group Methods for Particle Physics: Formal Analogies between Classical Statistical Mechanics and Quantum Field Theory', in *Synthese* 197: 3027–63.

Fraser, D. (2022), 'Particles in Quantum Field Theory', in E. Knox and A. Wilson (eds.), *Routledge Companion to Philosophy of Physics* (Routledge).

Fraser, D., and Koberinski, A. (2016), 'The Higgs Mechanism and Superconductivity: A Case Study of Formal Analogies', *Studies in the History and Philosophy of Modern Physics* 55: 72–91.

Guo, B. (2023), 'Next Best Thing—What Can Quantum Mechanics Tell Us About the Fundamental Ontology of the World?'. URL https://philsci-archive.pitt.edu/22673/.

Haag, R., Hugenholtz, N. M., and Winnink, M. (1967), 'On the Equilibrium States in Quantum Statistical Mechanics', in *Communications in Mathematical Physics* 5: 215–36.

Hartmann, S. (1998), 'Idealization in Quantum Field Theory', in *Poznan Studies in the Philosophy of the Sciences and the Humanities* 63: 99–122.

Jansson, L., and Saatsi, J. (2019), 'Explanatory Abstractions', in *The British Journal for the Philosophy of Science* 70: 817–44.

Koberinski, A. (2019), 'Parity Violation in Weak Interactions: How Experiment Can Shape a Theoretical Framework', in *Studies in History and Philosophy of Modern Physics* 67: 64–77.

Koberinski, A., and Fraser, D. (2023), 'Renormalization Group Methods and the Epistemology of Effective Field Theories', in *Studies in History and Philosophy of Science* 98: 14–28.

Morrison, M. (2015), *Reconstructing Reality: Models, Mathematics, and Simulations* (New York: Oxford University Press).

Morrison, M., and Morgan, M. S. (1999), 'Models as Mediating Instruments', in M. S. Morgan, and M. Morrison (eds.), *Models as Mediators* (Cambridge: Cambridge University Press), 10–37.

Reutlinger, A., and Andersen, H. (2016), 'Abstract versus Causal Explanations?', in *International Studies in the Philosophy of Science* 30: 129–46.

Ruetsche, L. (2011), *Interpreting Quantum Theories* (New York: Oxford University Press).

Wallace, D. (2001), 'The Emergence of Particles from Bosonic Quantum Field Theory'. arXiv:0907.5294v1.

Wallace, D. (2018), 'Lessons from Realistic Physics for the Metaphysics of Quantum Theory', in *Synthese* 197: 4303–18.

Wallace, D. (2020), 'On the Plurality of Quantum Theories: Quantum Theory as a Framework, and its Implications for the Quantum Measurement Problem', in S. French, and J. Saatsi (eds.), *Scientific Realism and the Quantum* (Oxford: Oxford University Press), 78–102.

II

CONCEPTS OF OPEN AND CLOSED SYSTEMS IN AND BEYOND PHYSICS

6
Newtonian Research and the Open Systems View

William L. Harper

6.1 Introduction

On the closed systems view, isolated systems are taken as fundamental and when a system is under the influence of its environment this is described in terms of a coupling between it and a separate system, which taken together are isolated. Michael Cuffaro and Stephan Hartmann (2021) argue against this view, in favour of an alternate open systems view, for which systems interacting with their environment are conceived of as fundamental. Cuffaro and Hartmann point out (Cuffaro and Hartmann 2021: 2) that Newton's consideration of the possibility of divine influence to prevent the cosmos from descending into chaos (Newton 1979: 402) showed that, in contrast to Laplace, he did not take his account of the solar system as a closed system.[1] They take Laplace's view, on which the ideal of natural philosophy is to develop laws and details that would account for the whole world as a pre-established order, as a salient example of the closed systems view they oppose.

This chapter will explore the extent to which the basic approach to explain solar system motions, in the research extending applications of Newton's theory, is an open systems methodology. The contrast between Newton's treatment of the sun's action on Jupiter's moons and his less successful treatment of its action on our moon will illuminate the importance of the opening-up of orbital systems to Newton's practice and methodology. Although Clairaut's solution to the lunar precession problem is a deeper understanding of the mathematical details of a specific three-body model of interaction among the moon, earth, and sun; the incorporation of details leading to measuring the mass of the moon and the successful application to understanding the precession of the equinoxes result from the opening-up of factors to be included in that model. These developments, as well as those resulting in empirically more accurate models, on to the Hill–Brown theory of the motion of our moon, are examples of successes obtained by opening up models to include additional factors.[2]

I take the search for additional factors to be included in a model being developed as an open systems view treatment of that model.

William L. Harper, *Newtonian Research and the Open Systems View*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press.
© Oxford University Press (2026). DOI: 10.1093/9780198929277.003.0007

We will see that additions of factors of many different sorts all count as examples of openings-up of models in research developing applications of Newton's theory. The empirical results Newton cites to support the application of his Rule 3 to accept his theory of universal gravitation are successfully developed in the applications by Clairaut, d'Alembert, and Euler to our moon's motion and the precession of the equinoxes. The open systems aspects of these developments are also exemplified in Clairaut's calculation of the return of Halley's comet. Newton's Rule 4, and the empirical successes that inform it, will help us to understand how open systems methodology is exemplified in later work applying Newton's theory in solar system research.

We will consider Newton's surprising solution to the two-chief-world-systems problem. Kepler's sun-centred model and Tycho's earth-centred model of our solar system motions are both wrong, but the Kepler model is an idealisation that better approximates the true motions and better affords resources for empirically finding successful corrections to apply. What Newton counts as the true motions are what would count as a sufficiently corrected model that could be appropriately accepted as an isolated system. We will see that this closed-system characterisation of what Newton would count as true motions in our solar system is an application of corollary 5 of his laws of motion. We will also see that corollary 6 of Newton's laws of motion opens this up to consider additional forces acting on the bodies in that system. So long as these additional forces are sufficiently close to being equal and parallel, they will recover the same empirical results as the idealised isolated-system model. In an unpublished document, Newton speculates about hypothetical extra forces that would transform our solar system into a Tychonic alternative. We will consider how, on Newton's methodology, one might respond to this open systems speculation.

The above mentioned results by Clairaut and d'Alembert suggest that the open systems nature of the development of empirical corrections of models can offset the closed-system nature of what Newton would count as the true motions being searched for. We will expand on this by considering solar-system research following Laplace's solution to the Jupiter–Saturn great inequality. The discovery of Neptune exemplifies the open systems nature of research applying Newton's theory to our solar system. We will also consider the Mercury perihelion problem that helped lead to the transition to Einstein's more fundamental theory.

6.2 Newton on Jupiter's Moons

Newton notes that the orbits of Jupiter's moons do not differ sensibly from uniform motion on concentric circles about the centre of Jupiter (Newton 1999: 797). He cites as phenomenon 1, in book 3 of his *Principia*, that these satellites of Jupiter

satisfy the area rule and the harmonic rule (ibid.). The data available accurately fit this idealised isolated-system model.[3]

Newton's basic argument for universal gravitation opens with his inference from this phenomenon 1 to the centripetal direction and inverse-square variation of the force maintaining those moons in their orbits of Jupiter (Newton 1999: 802). He cites prop. 2 or prop. 3, book 1, for the centripetal direction (ibid.). Corollary 1 of prop. 2 affords systematic dependencies that make an area rule orbit measure the centripetal direction of the force maintaining a body in that orbit, modelled as an isolated system (Newton 1999: 447; Harper 2010: 24–6).

Proposition 3, book 1, extends this inference to compounded forces. It argues that a constant areal rate shows that the remaining force by which the orbiting body is urged, after subtracting the compounding forces, will be directed towards the central body (Newton 1999: 448) Given that the data accurately fit this idealised isolated-system model, this proposition directly supports applying this inference to an extended model which includes the action of the sun on the system of Jupiter and its moons.

Later in his argument, Newton appeals to the regular motion of Jupiter's moons to support the proportionality of the weights of Jupiter and its satellites towards the sun to the quantities of their matter (Newton 1999: 807). The regularities in the orbits show the near equalities of the accelerations of the moons and Jupiter towards the sun.[4] The empirical regularities of data on Jupiter's moons, which accurately fit an idealised isolated-system model of orbits about Jupiter, afford significant evidence about how to open that system up to include its orbit of the sun.

6.3 Newton on Our Earth's Moon

Unlike the case of Jupiter's moons, the available data on our moon's motion included empirical deviations from area law motion. Newton points out that his inference from the area law motion to the centripetal direction of the force maintaining our moon in its orbit ignores these empirical deviations from the area rule motion generated by the action of the sun (Newton 1999: 801). He successfully treated the variational inequality discovered by Tycho (Wilson 2001: 141) by giving an accurate account of the violations of the area rule, exhibited by this inequality, as due to the action of the sun on the earth–moon system (ibid., 152–3).

In the first part of proposition 3, book 3, Newton states that it is evident from this area rule phenomenon (and prop. 2 or prop. 3 book 1) that the force maintaining the moon in its orbit of the earth is directed towards the earth (Newton 1999: 802). His successful treatment of the variational inequality supports this inference. The area rule holds, to good approximation, and the deviations can be ignored because

they are due to the action of the sun. In the second part of prop. 3, book 3, he argues that this force maintaining the moon in its orbit is inversely as the square of the distance of its places from the centre of the earth from the very slow motion of the moon's apogee (ibid.). He goes on to claim that this motion arises from the action of the sun and is to be ignored in this basic argument (Newton 1999: 803). The inverse-square law holds to good approximation, and the deviations can be ignored because he will show that they are due to the action of the sun. The claim that he will show that these deviations are due to the action of the sun, however, is incorrect. Newton's treatment only accounts for half of the lunar precession (Waff 1995: 35).

In his argument for prop. 4, book 3, Newton appeals to the moon-test agreement between inverse-square adjusting the centripetal acceleration exhibited by the moon's orbit, and the acceleration of terrestrial bodies measured by seconds pendulums, to infer that the force maintaining the moon in its orbit is terrestrial gravity (Newton 1999: 804). In his scholium to prop. 4, Newton considers a hypothetical system of several moons around the earth with the lowest one small and nearly touching the tops of the highest mountains (Newton 1999: 805). He applies the harmonic rule, which Kepler discovered for the planets, to infer that the force maintaining those moons in their orbits and terrestrial gravity would be similar to one another and equal, so that his first two rules for reasoning would require that they have the same cause (ibid.).[5] He goes on to argue that this lowest little moon must either be lacking in gravity or else fall twice as fast as heavy bodies generally do (ibid.). This scholium argument (like Newton's appeals to prop. 3, book 1, for Jupiter's moons) constrains what would be an acceptable opening-up of a model for motions of a moon. This lowest little moon would be subject to the acceleration of terrestrial gravity. If it were another force that maintained it in its orbit then it would be accelerated by both forces at once.

6.4 Later Work Applying Newton's Theory to Our Moon

We have seen that Newton's treatment of the lunar precession only accounts for half of it. This failure to account for the extra lunar precession extended into the initial treatments by Clairaut, Euler, and d'Alembert (Waff 1995: 36–40). These treatments were by differential equations in which some higher-order terms were ignored (ibid.). In his memoir of January 1749, Clairaut reported that when he labouriously recalculated taking these terms into account, he found that this successfully accounted for the extra precession (Waff 1995, 44–6). This is Clairaut's success at recovering the appropriate results for treating the accepted three-body isolated system corresponding to the earth, moon, and sun.

In 1749,[6] d'Alembert was able to use Bradley's discovery of the 18-year periodic nutation of the earth to measure the mass of the moon (Wilson 1995: 51). He was

able to apply that empirical result to calculate the mass ratio of the moon to the earth as 1/80 (ibid.). The present day value of this ratio is about 1/81.3.[7] Newton gives 1/39.788 as this ratio (Newton 1999: 879), which he calculated from attempting to compare the force of the moon to move the sea to the force of the sun to move the sea (Newton 1999: 875–87). D'Alembert's measurement is an enormous improvement over Newton's.[8] He achieved it through successfully opening up the model of the bodies being investigated.

The mass of the moon d'Alembert inferred from the nutation was further supported by his application of it to derive the, long accepted, 26,000-year periodic precession of the equinoxes (Smith 2014: 292). Euler developed an improved mathematical treatment of the mechanics of rigid bodies, which recovered the measured mass of the moon and its application to explain the precession of the equinoxes (Wilson 1995: 52–4; Wilson 1987). These results, by d'Alembert and Euler, were recovered from adding details about the shape of the earth, to what should count as an appropriate model of the earth, moon, and sun. This counts as a further opening-up of factors to be included in the model of that three-body system.

The Hill–Brown lunar theory was constructed in the years from 1877 to 1908 and adopted as the basis for lunar ephemerides of the US, UK, Germany, France, and Spain beginning in 1923 (Wilson 2010: vii). It was the most accurate lunar theory that had ever been constructed. George Smith indicates that it contained over fourteen hundred separate perturbational terms derived from gravity theory (Smith 2014: 302–3). The enormous number of such terms suggests that the Hill–Brown theory was developed as an open systems account of the lunar motions, in the course of which many additional factors were added to generate empirically supported refinements.

In addition to these terms generated from gravity, the theory had another term needed to bring calculations into agreement with observation. Brown called this 'the great empirical term' (Smith 2014: 303), and raised the question of whether its source is a fluctuation in the rotation of the earth or instead some further unaccounted for force acting on the moon (ibid.). It turns out that one part of the tidal effect of the moon involves an acceleration in the motion of the moon, while the other part arises from a slowing of the rotation of the earth (Smith 2014: 30). In 1939 Spencer Jones showed that that there was, indeed, a slowing in the rotation of the earth (Smith 2014: 304; Wilson 2010: 385–6). This slowing of the earth's rotation makes for a systematic error in sidereal time. In 1952 the International Astronomical Union (IAU) adopted Ephemeris time, based on the length of the tropical year for 1900 (Seidelmann et al. 1992: 79). In 1985 the IAU adopted atomic time (ibid.: 85).

These results, generated from the development of the Hill–Brown lunar theory, are examples of treating that theory as an open system account of the motion of our moon.

6.5 Newton's Rules for the Study of Natural Philosophy

Newton cites four rules of reasoning for doing natural philosophy. Here are the first two[9]:

> **Rule 1:**
> *No more causes of natural things should be admitted than are both true and sufficient to explain their phenomena.* (Newton 1999: 794)

> **Rule 2:**
> *Therefore, the causes assigned to natural effects of the same kind must be, so far as possible, the same.* (Newton 1999: 795)

We have seen that Newton appeals to these two rules to back up his inference to identify the force maintaining the moon in its orbit as terrestrial gravity. We will see that the injunction to avoid assigning superfluous causes endorsed by these rules can help defend the Newtonian solution to the two-chief-world-systems problem, without requiring that our solar system be counted as a closed system.

Newton's third rule expresses his commitment to inductive inference:[10]

> **Rule 3:**
> *Those qualities of bodies that cannot be intended and remitted [i.e., qualities that cannot be increased and diminished] and that belong to all bodies on which experiments can be made should be regarded as qualities of all bodies universally.* (Newton 1999: 795)

His discussion of this rule includes the following application of it to his theory of universal gravitation:

> Finally, if it is universally established by experiments and astronomical observations that all bodies on or near the earth gravitate [*lit.* are heavy] toward the earth, and do so in proportion to the quantity of matter in each body, and that the moon gravitates [is heavy] toward the earth in proportion to the quantity of its matter, and that our sea in turn gravitates [is heavy] toward the moon, and that all planets gravitate [are heavy] toward one another, and that there is a similar gravity [heaviness] of comets toward the sun, it will have to be concluded by this third rule that all bodies gravitate toward one another. (Newton 1999: 796)

We have noted (some of) the strong evidence Newton was able to appeal to in supporting his inference to accept gravitation towards the earth, proportional to their quantities of matter, for terrestrial bodies and the moon. We have noted Clairaut's successful treatment of the lunar precession, which may be counted as just a better understanding of a given three-body system. We have, also, noted

d'Alembert's successful measurement of the mass of the moon, which greatly improved upon Newton's attempt to estimate it from the role of the moon in accounting for the tides.[11] In addition, we have noted the successful treatments of the precession of the equinoxes by d'Alembert and Euler. As we have seen, these additional results are openings-up of the models being developed.

Newton was able to include Halley's basic orbital elements of the retrograde orbit of the comet we now know as Halley's comet, in the third edition of his *Principia* (Newton 1999: 934–6). The return of this comet in 1759, was another significant success for Newton's theory (Waff, 1995a: 79–82). Clairaut's calculation of it involved calculations of the perturbations induced by the attraction towards Jupiter, and Jupiter's action to change the sun's place, as well as the attraction of the comet to Saturn (Wilson 1995a: 83). On the basis of his calculation Clairaut had predicted a perihelion time of mid-April, give or take a month (Wilson, 1995a: 85). After the comet was successfully recovered, astronomers were able to empirically determine the perihelion to have occurred on 13 March (ibid.). Wilson points out that later developments show that perturbations due to planets other than Jupiter and Saturn caused the 1759 return of that comet to be 8 days earlier than it would otherwise have been (Wilson 1995a: 86). The increased accuracy afforded by these additional factors exhibits the open systems nature of the historical development of successful application of Newton's theory to understand the motion of that comet.

Accepting the foregoing application of Rule 3 would count Newton's theory of universal gravitation as a proposition gathered from phenomena by induction. This sets the stage for considering Newton's fourth rule of reasoning:[12]

> **Rule 4:**
> *In experimental philosophy, propositions gathered from phenomena by induction should be considered either exactly or very nearly true notwithstanding any contrary hypotheses, until yet other phenomena make such propositions either more exact or liable to exceptions.* (Newton 1999: 796)

This fourth rule of reasoning by Newton expresses an illuminating characterisation of central features of theory acceptance in scientific practice today. One important feature is that the acceptance endorsed is explicitly provisional. Further phenomena can make the inferred propositions liable to exceptions, as well as making them more exact. Another important feature is that the inferred propositions can be accepted as approximations. The most important feature is the answer to the question:

> What distinguishes what ought to be counted as 'propositions gathered from phenomena by induction' from propositions to be dismissed as mere 'contrary hypotheses'?

The basic answer is empirical success:

> A mere contrary hypothesis is a proposition that does not realise sufficient empirical success to count as a serious rival alternative.

A central advantage of the Newtonian framework is that applications of it have afforded empirical successes that go beyond the accurate prediction of empirical phenomena, which is the focus of the hypothetico-deductive account of scientific inference. Newton's basic inferences to inverse-square centripetal forces don't just show that those forces predict the orbital phenomena from which they are inferred. They are backed up by systematic dependencies, generated from the laws of motion, which make those phenomena accurately measure the centripetal direction and inverse-square variation of the accelerations which explain them.[13]

George Smith's account of the extraordinary success achieved in the Newtonian research programme is a striking example of yet another kind of empirical success:

> Those engaged in the 'normal science' of gravity research after Newton were not merely testing his theory or tightening the bounds of accuracy to which it was known to hold. In their efforts to represent the true motions ever more closely in a sequence of successive approximations, they were discovering which physical details make a difference in those motions and what differences those details make. (Smith 2014: 284)

This sequence, which produced successively more accurate approximations, was a sequence in which more and more physical details that make a difference were discovered, and more and more knowledge of what differences those details make was discovered. It was a sequence of empirical successes that counted as a sequence of increasingly high standards that any rival to Newton's theory would have to meet to avoid being legitimately dismissed as a mere contrary hypothesis, according to Rule 4.

As we will see, the open systems treatment of motions of bodies played a significant role in the extraordinarily successful research applying Newton's theory to the motion of planets in our solar system.

6.6 Newton on Solar System Research

The following passage, from an intermediate draft between the initial De Motu tract and the *Principia*, indicates Newton's appreciation of the daunting complexity of the motions of the planets in our solar system:

> There are as many orbits of a planet as it has revolutions, as in the motion of the moon, and the orbit of any one planet depends on the combined motion of all

> the planets, not to mention the action of all these on each other. But to consider simultaneously all these causes of motion and to define all these motions by exact laws admitting of easy calculation exceeds, if I am not mistaken, the force of any human mind. (Wilson 1989: 253)

This passage was not included in the published *Principia* (Smith 2014: 269).

Newton was able to use his theory of gravity to measure the masses of the sun and of the planets with satellite orbits (Newton 1999: 813). This allowed him to argue that, given the hypothesis that the solar system can be treated as an isolated system,

> *The sun is engaged in continual motion but never recedes far from the center of gravity of all the planets.* (Proposition 12, Book 3, Newton 1999: 816).

This, in turn, enabled him to apply Corollary 4 of his laws of motion:

> *The common center of gravity of two or more bodies does not change its state whether of motion or of rest as a result of the actions of the bodies upon one another; and therefore the common center of gravity of all bodies acting upon one another (excluding external actions and impediments) either is at rest or moves uniformly straight forward* (Newton 1999: 421)

to generate his surprising resolution of the two-chief-world-systems problem. The earth-centred and sun-centred systems are both wrong; but the sun-centred system can be a useful initial idealised model.

In prop. 13, book 3, Newton argues:

> ... it follows that if the sun were at rest and the remaining planets did not act upon one another, their orbits would be elliptical, having the sun at their common focus, and they would describe areas proportional to the times. (Newton 1999: 817–8)

These Kepler orbits can be accepted as idealisations that correspond to the motions the planets would have if the only force acting on each were its weight towards the sun. As George Smith has illuminatingly put it:

> Newtonian idealizations are special precisely because the theory, taken exactly, implies that systemic discrepancies between calculation and observation must be owing *only* to factors in the world not taken into account in the calculations. (Smith 2014: 277)

The solar system research applying Newton's theory to successively revise idealised models of the motion of a planet to make the models more accurate, by taking into

account additional factors in the world, is very much an open systems treatment of the systems of motions being modelled.

Consider the basic *n*-body Newtonian point-mass model for computing the motions of planets in our solar system. This solution to the problem of distinguishing what Newton (1999: 412) would count as the true motions (underlying the phenomena exhibited by the empirically accessible relations among those bodies) is a representation of the sun and planets as an isolated system. No forces are considered except the gravitational interactions of those bodies among themselves.

Stein's classic treatment of Newtonian space-time takes corollary 5 of the laws of motion as fixing the relativity with which the motions among themselves of a group of interacting bodies are fixed by a Newtonian model of the interactions among them (Stein 1967: 182).

> **Corollary 5:**
> *When bodies are enclosed in a given space, their motions in relation to one another are the same whether the space is at rest or whether it is moving uniformly straight forward without circular motion.* (Newton 1999: 423).

The interacting bodies are treated as an isolated system. No forces are considered beyond the interactions among the bodies in it.

Newton's corollary 6 opens this up to consider a special class of additional forces acting on the bodies in the system.

> **Corollary 6:**
> *If bodies are moving in any way whatsoever with respect to one another and are urged by equal accelerative forces along parallel lines, they will all continue to move with respect to one another in the same way as they would if they were not acted upon by those forces.* (ibid.)

Considering these hypothetical additional accelerative forces is an open systems treatment of the bodies being modelled.

At the 2014 meeting of the Philosophy of Science Association, Smith presented a paper titled 'Newtonian Relativity: A Neglected Manuscript, an Understressed Corollary'. Corollary 6 was what he characterised as understressed. The manuscript was a passage Newton wrote to be added to the end of the second edition of his *Principia*. Here is Smith's translation of what it says:

> *In order for the Earth to be at rest in the center of the system of the Sun, Planets, and Comets, there is required both universal gravity and another force in addition that acts on all bodies equally according to the quantity of matter in each of them*

> *and is equal and opposite to the accelerative gravity with which the Earth tends to the Sun, tending along parallel lines on the same flat surface with the line drawn from the center of the Sun to the center of the Earth.*
>
> For, such a force, acting on all bodies equally and along parallel lines, does not change their position among themselves, and permits bodies to move among themselves through the force of universal gravity in the same way as if it were not acting on them. Since this force is equal and opposite to its gravity toward the Sun, the Earth can truly remain in equilibrium between these two forces and be at rest. And thus celestial bodies can move around the Earth at rest, as in the Tychonic system. (Smith 2014a)

Newton did not include this passage when the second edition of his *Principia* was published in 1713 (Smith 2014a).

We can think of this speculative Tychonic alternative to his solar system model as one showing that this specific opening-up of that model would recover all the relative motions and forces of that model. This application of corollary 6 showed that adding these equal and parallel accelerations would not result in any empirically supported changes in the corresponding isolated-system model of the sun and planets. It would, however, change what would be accepted as the true motions among the sun and planets.

I think these speculative additional forces would be dismissed, as a mere contrary hypothesis, according to Newton's rules for doing natural philosophy. This alternative hypothesis would explain exactly the same solar-system phenomena by admitting all the same causes represented by the gravitational forces in the accepted Newtonian model. The hypothesised extra forces would not explain any phenomena that were not already explained by that accepted model. This suggests Newton's Rule 1 would endorse dismissing their introduction to generate an alternative model as an introduction of superfluous causes. Further, unlike the Newtonian gravitational interactions among solar system bodies, these extra forces would not be pairwise interactions among bodies. This suggests that they might also be dismissed by Rule 2, unless some other body were included in the speculation.[14]

Newtonian attractions towards distant stars would be expected to be sufficiently close to equal and parallel that they would not result in any empirical changes to the solar system model. Unlike this Tychonic speculation they are not contrary hypotheses. They don't change what count as the true motions in our solar system.

For Tycho Brahe, empirical failure to detect annual stellar parallax was taken to show that the earth does not orbit the sun (Thoren 1989). For Newton it would indicate that the stars were farther away than the distances that would have, at his time, afforded empirically detectable annual parallax.[15]

6.7 Later solar system research

Let us begin with Laplace's solution to the Jupiter–Saturn great inequality. Certain sizeable vagaries in the mean motions of Jupiter and Saturn had been recognised by Kepler in 1625 and continued to resist attempts by astronomers to account for them long after the acceptance of Newton's theory (Wilson 1985: 20). In 1785 Laplace announced that he had shown that these anomalies in the motions of Jupiter and Saturn could be accounted for on the assumption of inverse-square gravitation (Wilson 1985: 22). This result transformed the confusing data of these deviations of motions of Jupiter and Saturn from their basic Kepler orbits into data fitting a theory-mediated phenomenon.

The application of the theory characterising its physical source directly resulted in discovering this long-term pattern, which was then accepted as holding for data relating the orbital motions of Jupiter and Saturn.[16] Like Clairaut's solution of the lunar precession problem, Laplace's solution to the Jupiter–Saturn great inequality resulted from taking into account higher-order terms in a three-body model being investigated (Wilson 1985: 23–4, 256–77, 286–90). His successful account of the Jupiter–Saturn great inequality was a characterisation of deviations of angular positions from their basic Kepler orbits, generated by their gravitational interactions (Wilson 1985: 34–6).

This can be thought of as just a deeper understanding of the isolated 3-body gravitational model of Jupiter, Saturn, and the sun. Indeed by accounting for these long-outstanding difficulties, Laplace opened a path towards the improvement of theories and tables of all the planets, which transformed solar system research (Wilson 1985: 287).[17] And his success in applying Newton's theory to generate a refined account of solar system motions[18] may well have suggested counting its solar system model as a closed system. His famous remarks about research leading to a deterministic theory of everything certainly suggest that he would endorse taking a sufficiently refined Newtonian model of our solar system as a closed system (Laplace 1995: 4).

However, later research applying Newton's theory to our solar system affords significant support for the open systems view. Consider the research sequence leading to the discovery of Neptune. The accepted model of motions of Uranus was what would be taken as its basic Kepler orbit, together with the perturbations of that orbital motion corresponding to its interactions with Jupiter and Saturn (Morando 1995a: 216; Laplace 1966, Vol. 3: xxi, 314–7). Deviations from this accepted model, that would have been masked by the larger deviations due to interactions with Jupiter and Saturn, were observed (Smith 2014: 280–1; Morando 1995a: 216–7). This second-order phenomenon was found to be due to interactions with the newly discovered planet, Neptune.[19] This is a an open system treatment of the motions of Uranus. More generally, research sequences in which deviations from an accepted model for the motion of a planet are second-order

phenomena, which are expected to be caused by further Newtonian interactions, are examples of treating the accepted models as open systems.

6.8 Mercury's Perihelion Precession

A salient example of a second-order phenomenon discovered in the research applying Newton's theory to our solar system is the famous 43-arc-seconds-per-century precession of Mercury's perihelion not explainable by Newtonian gravitation.[20] In 1915 Einstein showed that his general theory of relativity would account for this residual precession of Mercury (Einstein 1915). One important part of the Einstein solution to the Mercury perihelion problem was that, in addition to this famous 43 arc seconds per century, his theory would also recover the 531 arc seconds per century that were recovered by Newtonian perturbations.

The basic programme for computing motions of planets for the *Astronomical Almanac* is an *n*-body Newtonian point-mass model with relativistic corrections (Seidelmann 1992: 281–2). Without the relativistic corrections that model would only account for those 531 arc seconds per century of Mercury's perihelion precession. The relativistic corrections introduce the empirical features needed to make the model correspond to an account of what these motions are according to General Relativity. This makes those features generate the full empirically determined precession of Mercury.

Einstein's static weak field limit condition insured that his theory of general relativity would recover dependencies corresponding to the Newtonian contributions of these planets to Mercury's precession (Smith 2014: 314). Our solar-system Newtonian empirical successes are all ones that are recovered by General Relativity (Smith 2014: 317–9).[21]

Endnotes

1. More generally (see Harper 2010: 386–8), Newton's practice and methodological remarks do not support that he is guilty of the hubris suggested by the impressive figure in Blake's portrait of Newton, who may consider that the diagram he is looking at gives him a God-like access to the world from the outside. For Newton, the investigator is inside the world, and always open to finding out more about it.
2. We will see that these factors came to be understood as including a slowing of the rotation of the earth, which led to the need for a new astronomical time standard.
3. The regularities exhibited in the periods and eclipses of the orbits of Jupiter's moon Io were sufficiently well established by 1676 that Rømer was able to use them to empirically calculate that light has a finite speed (Débarbat and Wilson 1989: 152–4); see also Cohen (1940).
4. Newton cites, without details, a calculation that is purported to show that a difference in accelerative gravities of Jupiter and a satellite towards the sun would result in a noticeable polarisation of the orbit of that satellite. We can think of the empirical bound on absence of orbital polarisation as yielding a bound on an Eötvös ratio $\Delta, 2(a_1 - a_2)/(a_1 + a_2)$, representing differences between the acceleration a_1 of Jupiter towards the sun and the acceleration a_2 of its satellite towards the

sun (Harper 2010: 278). Newton's data, for Jupiter's satellite Callisto, make a correct calculation yield a bound limiting such differences to $\Delta < 0.02$ (Harper 2010: 282–3).

5. See below for Rules 1 and 2.
6. Bradley's empirical result was published in 1748 (Wilson 1995: 47). D'Alembert heard an abstract read that summer (ibid.).
7. Seidelmann (1992: 706), puts the mass of the moon at 0.073483×10^{24} kg and the mass of the earth at 5.9742×10^{24} kg. This yields that the moon's mass is about 1/81.297 of the mass of the earth.
8. D'Alembert took his successful treatment of nutation, as due to the gravitation of the earth towards the moon, as affording strong evidence for Newton's account of gravity's reciprocal attraction between spatially separated bodies (Wilson 1995: 48).
9. These first two Rules were introduced in the first edition of his *Principia in 1687*, where he called them 'hypotheses'. In the second edition of 1713 he called them 'Rules' (Newton 1999: 198–200, 794).
10. This Rule was introduced in the second edition of 1713 (Newton 1999: 199). See Domski (2022) for an illuminating discussion of Newton's application of this Rule.
11. For Newton's work on the tides see Cartwright (1999: 35–44). For Laplace's more successful account see Cartwright (1999: 68–87).
12. This Rule was introduced in the third edition (Newton 1999: 2000). It is, however, anticipated in Newton's response to a challenge by Cotes, the editor of his second edition (Harper 2010: 347, 361–2).
13. Harper (2010: 24–8). These examples of empirical success afford stronger evidence supporting those inferences than would be supported by just showing that assuming those inverse-square accelerations would predict those phenomena (ibid.: 373–8). See also Harper (2020).
14. Smith's manuscript includes a calculation that having departures from being equal and parallel has no greater effect on the motion of Saturn than the departures from equal and parallel of the action of the sun on the outermost satellites of Jupiter. As we have seen, the effects of such departures were undetectable in the data available to Newton. The result of Smith's calculation is that our solar system would have to be orbiting a massive object nearly 4000 astronomical units from the earth with a mass of about 60 billion solar masses (Smith 2014a: 11).
15. In 1838 the empirical detections of stellar parallax were successfully achieved for three separate stars by Struve, Bessel, and Henderson (Dyson 1915: 18–9).
16. This long-term pattern has a period of about 900 years (Wilson 1985: 30). Jupiter passes Saturn about every twenty years (ibid.: 28). For about half of the period of the great inequality, each time Jupiter passes it, Saturn is slowed and Jupiter sped up; then, for the other half of that period, the opposite happens (ibid.: 33).
17. Wilson suggests that the unresolved large anomalies in the motions of Jupiter and Saturn had made the prospects for success of attempts to incorporate perturbations in the astronomical tables of planets seem unhopeful (ibid.).
18. Laplace was a leading figure refining the Newtonian model of solar system motions to take into account perturbations produced by gravitational interactions. In the third volume of his *Celestial Mechanics* (Laplace, 1966), which was published in 1802, Laplace's theory of Mercury included perturbations produced by the actions of Venus, the Earth and Jupiter (xix; 229–36). His theory of Venus included perturbations produced by the actions of the Earth, Mars, Jupiter, and Saturn (xix; 237–43). His theory of the motion of the Earth included perturbations by Venus, Mars, Jupiter, and Saturn (xx; 244–67). His theory of Mars included perturbations by Venus, the Earth, Jupiter, and Saturn (xx; 268–74). His theory of Jupiter included inequalities produced by the Earth, Saturn, and Uranus, but chiefly Saturn (xx; 275–98). His theory of Saturn included inequalities produced by Jupiter and Uranus (xxi; 299–313). His theory of Uranus included inequalities produced by the action of Jupiter and Saturn (xxi; 314–7).
19. Le Verrier and Adams had independently proposed that the source of these deviations, from the accepted model of the motion of Uranus, was gravitational attraction towards another planet (Morando 1995a: 218). In 1846, Galle, of the Berlin Observatory, successfully observed this

planet (ibid.: 219). Further observations, including a discovery of its satellite Triton, made it possible to more accurately determine its orbit, and mass. (ibid.: 221).

20. For more detailed accounts see Earman et al. (1993) and Roseveare (1982).
21. For the earth–moon interaction slowing the earth, we need to go beyond point-mass models; but, here again, we take general relativity to endorse whatever would be recovered by the appropriate Newtonian model (Smith 2014: 319). The fact that we don't include relativistic factors in constructing these models in no way undercuts the fact that general relativity is the more fundamental theory.

Bibliography

Biener, Z. and Schliesser, E. (eds.) (2014), *Newton and Empiricism* (New York: Oxford University Press).

Cartwright, D. E. (1999), *Tides: A Scientific History* (Cambridge University Press).

Cohen, I. B. (1940), 'Roemer and the First Determination of the Velocity of Light', in *Isis* 31: 325–79.

Cuffaro, M. E. and Hartmann, S. (2021), 'The Open Systems View'. arXiv:2112.11095v2.

Débarbat, S. and Wilson, C. (1989), 'The Galilean Satellites of Jupiter from Galileo to Cassini, Römer and Bradley', in Taton and Wilson (eds.) *Planetary Astronomy from the Renaissance to the Rise of Astrophysics. Part B: The Eighteenth and Nineteenth Centuries* (Cambridge: Cambridge University Press), 144–57.

Domski, M. (2022), *Newton's Third Rule and the Experimental Argument for Universal Gravity* (Routledge).

Dyson, F. W. (1915), 'Measurements of the Distances of the Stars', in *Science* 42: 13–22.

Earman, J., Janssen, M., and Norton, J. D. (1993), 'Einstein's Explanation of the Motion of Mercury's Perihelion', in J. Earman, M. Janssen, and J. D. Norton (eds.), *The Attraction of Gravitation: New Studies in the History of General Relativity* (Boston: Birkhäuser).

Einstein, A. (1915), 'Explanation of the Perihelion Motion of Mercury by Means of the General Theory of Relativity', trans. B. Doyle, in K. R. Lang and O. Gingerich (eds.), *A Source Book in Astronomy and Astrophysics, 1900–1975* (Harvard University Press), 820–5.

Harper, W. L. (2010), *Isaac Newton's Scientific Method: Turning Data into Evidence about Gravity and Cosmology* (Oxford University Press).

Harper, W. L. (2020), 'Newton's Scientific Method', in Schiesser and Smeenk (eds.), *The Oxford Handbook of Newton* (Oxford University Press).

Laplace, P. S. (1995), *A Philosophical Essay on Probabilities*, trans. F.W. Truscott and F.L. Emory (Toronto: Dover Publications).

Laplace, P. S. (1966), *Celestial Mechanics*, vols. 1–4, trans. N. Bowditch (New York: Chelsea Publishing Co.)

Morando, B. (1995a), 'The Golden Age of Celestial Mechanics', in Taton and Wilson (1995: 211–39).

Morando, B. (1995), 'Laplace', in Taton and Wilson (1995: 131–50).

Newton, I. (1979), *Opticks* (New York: Dover Publications, Inc.)

Newton, I. (1999), *The Principia, Mathematical Principles of Natural Philosophy: A New Translation*, trans. I. B. Cohen and A. Whitman (Berkeley: University of California Press).

Roseveare, N. T. (1982), *Mercury's Perihelion from Le Verrier to Einstein* (Oxford: Oxford University Press).

Seidelmann, P. K. (1992) (ed.), *Explanatory Supplement to the Astronomical Almanac* (Mill Valley: University Science Books).
Seidelmann, P. K., Guinot, B., and Doggett, L. E. (1992), 'Time', in Seidelmann (1992: 39–93).
Smith, G. E. (2014), 'Closing the Loop: Testing Newtonian Gravity, Then and Now', in Biener and Schliesser (2014: 262–351).
Smith, G. E. (2014a), 'Newtonian Relativity: A Neglected Manuscript, an Understressed Corollary', Manuscript.
Stein, H. (1967), 'Newtonian Space-Time', in *The Texas Quarterly* X: 174–200.
Taton, R. and Wilson, C. (eds.) (1989), *Planetary Astronomy from the Renaissance to the Rise of Astrophysics. Part A: Tycho Brahe to Newton* (Cambridge University Press).
Taton, R. and Wilson, C. (eds.) (1995) *Planetary Astronomy from the Renaissance to the Rise of Astrophysics. Part B: The Eighteenth and Nineteenth Centuries* (Cambridge University Press).
Thoren, V. E. (1989), 'Tycho Brahe', in Taton and Wilson (1989: 3–21).
Waff, C. B. (1995), 'Clairaut and the Motion of the Lunar Apse: The Inverse-Square Law Undergoes a Test', in Taton and Wilson (1995: 35–46).
Waff, C. B. (1995a), 'Predicting the Mid-Eighteenth-Century Return of Halley's Comet', in Tatom and Wilson (1995: 69–82).
Wilson, C. (1985), 'The Great Inequality of Jupiter and Saturn: From Kepler to Laplace', in *Archive for History of Exact Sciences* 33: 15–290.
Wilson, C. (1987), 'D'Alembert versus Euler on the Precession of the Equinoxes and the Mechanics of Rigid Bodies', in *Archive for History of Exact Sciences* 37: 233–73.
Wilson, C. (1989), 'The Newtonian Achievement in Astronomy', in Taton and Wilson (1989: 233–74).
Wilson, C. (1995a), 'Appendix: Clairaut's Calculation of the Comet's Return', in Taton and Wilson (1995: 83–6).
Wilson, C. (1995), 'The Precession of the Equinoxes from Newton to d'Alembert and Euler', in Taton and Wilson (1995: 47–54).
Wilson, C. (2001), 'Newton on the Moon's Variation and Apsidal Motion: The Need for a Newer "New Analysis"', in J. Z. Buchwald and I. B. Cohen (eds.), *Isaac Newton's Natural Philosophy* (MIT Press), 139–88.
Wilson, C. (2010), *The Hill–Brown Theory of the Moon's Motion: Its Coming-to-Be and Short-lived Ascendancy (1877–1984)* (New York: Springer).

7

Blackbody Radiation: The Open and Closed Systems Views and Complementary Reasoning Strategies

Molly Kao

7.1 Introduction

Philosophers have long been interested in questions related to the articulation of new scientific theories. Although the general consensus is that there is no infallible set of rules that one can follow in such contexts, several authors have drawn attention to fruitful methodological processes that are often used in theory development.[1] The practice of characterising such processes faces particular challenges when considering the early stages of an emerging theory, during which scientists are often beset by a wealth of conflicting information. Sometimes this comes in the form of observations that are inexplicable by leading extant theories, or it may be found in the form of assumptions that seem necessary to account for various empirical results but that are in tension with one another. It is thus important to consider such tricky cases to more fully understand how we can effectively develop a new theory.

Prior to the development of quantum theory at the dawn of the twentieth century, physicists generally accepted the idea that energy is correctly described as a quantity that can take on continuous values. However, in the late nineteenth century, various experiments began to point to the inadequacy of classical theories in accounting for certain observations. One of these was the phenomenon of blackbody radiation, which led Max Planck to introduce the notion of 'energy elements' in 1900, effectively proposing that energy was in fact not a uniquely continuous quantity. This was a groundbreaking hypothesis that eventually led to the articulation of the theory of quantum mechanics, one of physics' most successful theories to date. One curious feature of this episode, which has been much discussed, is the inconsistency of certain assumptions that Planck used in order to derive his theory. Despite this inconsistency, the theory was ultimately enormously fruitful.

The goal of this chapter is to conduct a historical study of the development of Planck's radiation theory, as well as the ensuing criticisms of this theory, in order to

Molly Kao, *Blackbody Radiation: The Open and Closed Systems Views and Complementary Reasoning Strategies*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press. © Oxford University Press (2026). DOI: 10.1093/9780198929277.003.0008

make a general argument about theory development in emerging contexts. Briefly, we can be led to interesting or fruitful results when we are able to mobilise complementary reasoning strategies. This is demonstrated in two different stages of the historical case study, each of which exemplifies this general claim.

First, in Section 7.2, I argue that Planck's treatment of the blackbody system reveals that the open and closed systems views (see the introduction to this volume) are both present in his methodology. In particular, the two conflicting and purportedly inconsistent assumptions present in Planck's initial radiation theory can be interpreted as two ways of conceptualising the relation between the blackbody and the radiation field. A statistical treatment of the energy corresponds to an open systems view, in which the blackbody and the exciting field are treated generally in terms of their interaction, whereas the thermodynamically influenced treatment of the blackbody's energy can be seen as corresponding to a closed systems view of the blackbody and its exciting field. The combination of these views thus contributed to a theoretical advance that would not have been otherwise possible.

Then, Section 7.3 provides an illustration of two different responses to the inconsistency by Planck's contemporaries, namely by James Jeans (1924) and Albert Einstein (1906/1989), which have not played a prominent role in philosophical discussions of the inconsistency to date. The primary goal here is to contrast their ways of reasoning given the information. A secondary aim in bringing up their responses is to make a historical point. I will argue that while there was certainly recognition of the problem of combining classical principles with a quantum hypothesis, what we actually see in Jeans and Einstein are criticisms that the assumptions lacked independent motivation, or that they led to predictions that were not borne out by observation. I will outline how Einstein worked with these assumptions slightly differently than Jeans did, using them as an indication of how to develop the theory further, demonstrating a broad heuristic view that is arguably characteristic of Einstein's reasoning in several domains. These responses show another way that complementary reasoning strategies allow us to pursue an emerging theory.

7.2 Planck and Blackbody Radiation

In this section, I will begin by presenting certain aspects of Planck's work on blackbody radiation in order to clarify how he arrived at a hypothesis of quantised energy, paying particular attention to the reasons why his assumptions are often regarded as being inconsistent.

Gustav Kirchhoff introduced the notion of a blackbody in 1860 in his work on the relation between the absorbing and emitting capacities of various bodies (Kirchhoff 1860). The previous year, he had shown that, for a body with emissive and absorptive capacities of e and a respectively, the ratio of e/a was a universal

function dependent on the wavelength of the energy being considered, and the temperature of the body. Building on this result, he introduced the idea of a theoretical body that absorbs all incident electromagnetic radiation, i.e., a blackbody with absorptive capacity $a = 1$. For such an object, since $a = 1$, the emission spectrum e would be a function of wavelength and temperature only.[2] In other words, the spectrum of radiation emitted by such a body would be independent of its composition. This would be an important point for Planck's later work.

In the years leading up to 1900, a number of observations had been made to determine the emission density of energy of various wavelengths at different temperatures, and several general features of the function describing this relation were known. For example, as the temperature increased, the maximum energy density was displaced towards shorter wavelengths λ.[3] For some time, the distribution law proposed by Wilhelm Wien looked as though it accurately described the full emission spectrum. Indeed, part of Planck's research programme was to derive a general form of this law, since he took it to be a correct description of the emission spectrum of a blackbody when he began his work on radiation. However, soon after, experiments conducted primarily by Lummer and Pringsheim (1900), began to show significant deviations from this expression, thus demonstrating that the Wien law was at best accurate for only a part of the spectrum. At this time, physicists recognised that, for a different range of frequencies in the energy distribution, the Rayleigh–Jeans law provided an accurate description in comparison to Wien's distribution law. Moreover, this law was shown by Lord Rayleigh to follow directly from principles of classical physics, thus providing it with a theoretical justification that Wien's law did not possess.[4] Thus, to summarise, in 1900, no single mathematical expression completely described the radiation density of a blackbody at all observed frequencies.

Nevertheless, Planck (1900a[1972]) was able to infer the form of the radiation law based largely on these two equations of limited validity, proposing the formula

$$u_\nu = \frac{A\nu^3}{e^{(B\nu/T)} - 1}, \tag{1}$$

where A and B are constants. This formula thus describes the density u of energy of frequency ν being emitted by a blackbody, for a particular temperature T. While Equation 1 agreed with all existing observations of the emission of radiation from a blackbody, it was obtained mainly through mathematical interpolations from the previously mentioned equations possessing limited validity in order to arrive at the desired empirical result. It was only after its proposal that Planck attempted to provide a theoretical interpretation of this law. To do so, Planck first calculated restrictions on the entropy of a blackbody in its equilibrium state in two different ways. He then compared and combined the two forms to obtain a more precise expression of the total entropy of the blackbody. Subsequent rearrangement yielded the energy distribution law sought for. These two ways of calculating the

entropy are often identified as being inconsistent. In the next section, I will present these two ways of calculating the entropy restrictions as corresponding to the application of the open and closed systems views. To begin, I will first give a brief overview of the relevant notion of open and closed systems.

7.2.1 Open and Closed Systems

According to Cuffaro and Hartmann (2024), 'a system is considered to be closed if it does not exchange energy, matter, heat, information, or anything else with its environment' (2024: 2). While it is arguable that no system is truly closed in every respect from a material point of view, formal representations of closed systems have been very successful. For my purposes, I will primarily be considering how the system in question should be modelled. We can, on the one hand, consider the effect of outside influences by coupling the system of interest to its environment, and considering the evolution of this larger system. This new, larger system is then in turn considered to be independent from its own environment, that is, the system of interest is represented as part of a closed system. On the other hand, the influence of the environment on an open system can be modelled by including it as part of the dynamical equations representing the evolution of the system of interest.

Cuffaro and Hartmann are primarily concerned with issues related to the fundamentality of open and closed systems. They present the notion of a 'view', which encapsulates certain commitments, namely, '(i) a set of methodological presuppositions in accordance with which we characterize the objects of a given domain, that is (ii) motivated by a particular metaphysical position concerning the nature of those objects' (2024: 5). They argue that adopting a particular view in quantum theory is part of what is implied by accepting a particular theoretical framework, and they argue subsequently for the fundamentality of a framework that is formulated in accordance with what they call the open systems view. In what follows, my purpose is to show, how on one way of understanding the notions of open and closed systems views, they may manifest, not necessarily as competing frameworks of a theory, but as non-competing ways of thinking about systems in a more limited context, particularly when it comes to the process of investigating new phenomena or pursuing new theories. To that end, I take the 'open and closed systems views' to characterise, more generally, ways of reasoning about a system that either explicitly take into consideration certain relevant aspects of its immediate environment, or consider a system in a more insular way. In particular, these approaches manifest in the context of investigating blackbody phenomena through different ways of describing the blackbody itself. For this reason, I also use the word 'approach' to characterise Planck's reasoning. This is because I do not take Planck to be committed to either of the two views; instead, I would argue that the combination of the two ways of reasoning implies being noncommittal, to a certain extent, about the 'best' way to model the system. This reconstruction

of Planck's reasoning thus provides a historical example of how a blend of open and closed systems views was fruitfully used in the development of a new theory.

7.2.2 Inconsistent Assumptions

As stated previously, the purported inconsistency in Planck's theory arises in the theoretical derivation of Equation 1, which required two different ways of calculating the entropy of a resonator. It is worth noting here that Planck was conceiving of the blackbody system in a very general way to perform his calculations. Specifically, he employed a model of a blackbody that was justified by Kirchhoff's Law, which, as previously mentioned, posited that the behaviour of radiation is independent of the specific properties of the system. Planck thus incorporated the idea of Hertzian resonators—small, nonresistive, oscillating electric circuits—to represent an enclosure interacting with the electromagnetic radiation. In such a model, the energy of the system is expressed in terms of the total energy of the resonators comprising the blackbody. Here, I consider the system under investigation to consist of the nonresisting oscillators, and the differences between the closed and the open systems views can be seen when considering the representation of the relation to the incident radiation or exciting field.

The first expression restricting the entropy of a resonator was the result of a rearrangement of an expression for the total entropy of the system in question. To obtain this expression, Planck was inspired by Boltzmann's earlier work on the treatment of the kinetic theory of gases in which the probability of a system's being in a particular macrostate is based on a calculation of the number of microstates giving rise to it. Planck thus supposed that the energy of the system could be divided into parts, at least numerically. Crucially, he made the assumption that the total energy of the system (U_N) could be considered as a collection of a number of elements (P), each with energy of quantity ϵ, where the size of ϵ was not yet determined. This step allowed him to calculate the number of microstates of the blackbody system, which was necessary for the calculation of total entropy. This process yields the first restriction on the entropy of a resonator in terms of its energy U.

In particular, Planck considered a large number N of identical resonators of frequency ν. Inspired by Boltzmann, Planck takes the total entropy S_N to be $k \log W + constant$, where W is the number of microstates corresponding to the system's possessing total energy U_N. The number of ways to divide the P energy elements between N resonators is given by $W = (N+P-1)!/P!(N-1)!$ By ignoring the 1s in relation to large Ns and Ps and applying Stirling's formula, we get

$$W = \frac{(N+P)^{N+P}}{P^P N^N}. \tag{2}$$

This result can be substituted into the definition of entropy, $S = k \log W$, and the resulting equation for the total entropy of the system is thus

$$S_N = k\left[(N+P)\log(N+P) - P\log P - N\log N\right]. \tag{3}$$

However, total entropy S_N is simply the entropy of a single resonator S, times the number of resonators N. Furthermore, the total energy of the resonators, U_N, is equal to $P \times \epsilon$, the number of energy elements times the amount of energy in each element, and U_N is also equal to NU, where U is the average energy of a resonator. Thus, P/N is equivalent to U/ϵ, and Equation 3 can be rewritten as

$$S = k\left[\left(1+\frac{U}{\epsilon}\right)\log\left(1+\frac{U}{\epsilon}\right) - \frac{U}{\epsilon}\log\frac{U}{\epsilon}\right], \tag{4}$$

where S is the entropy of a resonator, k is a constant, U is the average energy of a single resonator, and ϵ is an as-yet-unspecified amount of energy.

In this context, the blackbody resonators are not isolated, since the presence of the incident radiation provides energy to the blackbody. However, as mentioned, the calculation of the total energy U_N of the blackbody makes use of the concept of 'elements' of energy of size ϵ that are portioned out amongst the oscillators. Although these energy elements do not depend directly on the exciting field, the energy that is obtained from the incident radiation is implicitly included in the description of the blackbody's energy elements. We can thus consider this an example of the use of an open systems view, in which the blackbody, consisting of a number of resonators, is in constant interaction with the surrounding radiation field. Indeed, as Einstein states, 'one has to assume that there exists an arbitrarily small number of freely moving molecules besides the resonators, which by virtue of their collisions with the ions can transmit energy from resonator to resonator' (1906/1989: 194).

The second restriction on entropy relies heavily on Planck's earlier work in radiation theory, in which he calculated the relation between the radiation density u_ν and the average energy of a resonator U to be

$$u_\nu = \frac{8\pi\nu^2}{c^3}U. \tag{5}$$

He did this by using classical electromagnetic theory to determine the expressions for the absorption and emission of energy respectively, and equating them to obtain the relation that holds at the equilibrium state for an individual resonator. He then used this information to derive a form of Wien's Displacement Law, which states that the entropy must be a function of energy over frequency,

$$S = f\left(\frac{U}{\nu}\right), \tag{6}$$

which can be seen as a restriction on the form of any expression for entropy in the equilibrium state. To be more precise, Planck used his relation between a resonator's absorbed energy and the intensity of the incident radiation to calculate the relation between the average energy of a resonator and the intensity of a monochromatic polarised ray K. With this relation between K and U, along with the Kirchhoff–Clausius law, one obtains the form of u_ν in Equation 5.

It is important to note that in the calculation of this formula, the relation between the energy distribution and the energy of a single resonator is one that Planck had obtained based on Maxwell's theory of electromagnetism, which considers the behaviour of a resonator in a radiation bath. This theory explicitly gives the energy of a Hertzian oscillator as a function of the intensity of the exciting field. In other words, the exciting field, i.e., the environment with which the blackbody system is interacting, is represented in the description and evolution of the system. For this reason, we can characterise this approach to restricting the entropy equation as being consistent with a closed systems view. The exciting system is coupled explicitly to the resonators constituting the blackbody and the description of the system evolves accordingly.

The reason Equations 4 and 6 are significant is that they serve as restrictions on the form of entropy, whose combination allowed Planck to posit that the energy element ϵ must be proportional to frequency ν. This results in the following expression for the entropy of a resonator:

$$S = k\left[\left(1 + \frac{U}{h\nu}\right)\log\left(1 + \frac{U}{h\nu}\right) - \frac{U}{h\nu}\log\frac{U}{h\nu}\right]. \tag{7}$$

Here, h and k are universal constants. This particular expression then allowed Planck to calculate the energy density formula that was in accord with observations. Specifically, by differentiating Equation 7 with respect to U and setting the result equal to $1/T$, we obtain an expression for U which can then be substituted into Equation 5. We then obtain the final energy density formula,

$$u = \frac{8\pi h\nu^3}{c^3}\frac{1}{e^{h\nu/kT} - 1}. \tag{8}$$

I would note here that the foregoing presentation does not purport to give a faithful rendering of Planck's intellectual route to the arrival of the radiation law. Were that the case, the inclusion of these seemingly inconsistent assumptions would be puzzling, in addition to being problematic. However, this supports the idea that, at a methodological level, we can interpret elements of Planck's reasoning as incorporating both open and closed systems views. Badino (2015) makes

the case that Planck was working on several problems and approaches to the problem of radiation in parallel during a period of rapid developments, and thus the ensuing argument is not easily represented in terms of a straightforward argumentative line. It is perhaps particularly interesting to note that the more classical, Maxwellian approach corresponds to the closed systems view, which Cuffaro and Hartmann characterise as being considered more fundamental, historically. However, the incorporation of an open systems view led to a revolutionary change in the conception of the system, which in turn led to further theoretical developments. In this particular case then, a theoretical step forward was the result of overcoming the exclusive closed systems view inherent in the classical approach.

7.2.3 Discussion

It so happens that this is a particularly striking example of combining the open and closed systems views, insofar as a combination of the two was necessary for Planck to arrive at the accurate radiation theory. However, we can observe something similar in Planck's proposal of his second theory of radiation. Despite the success of his first radiation theory, Planck was never satisfied with its theoretical underpinnings. Planck wrestled with his own work for years, due to the fact that it was not his wish to revolutionise physics. Just as he had sought to maintain the absolute nature of the second law of thermodynamics in his earlier work, so he wished to resolve the tension between his two calculations of entropy in a way that would not overturn all of the classical theory. This led to his publication in 1912 of his second radiation theory. In this revised theory, Planck posited an asymmetry between the absorption and emission processes for the energy of the blackbody oscillators.[5] In this theory, Planck expressed the energy A absorbed by an oscillator up to a time t as

$$A = \frac{J_\nu}{4L} t, \tag{9}$$

where J_ν is the spectral component of the field intensity corresponding to the characteristic frequency ν of the oscillator, and L is a constant. In this way, the absorption process is described explicitly in terms of the exciting field, which is reminiscent of a closed systems approach. However, the emission process in this theory was treated statistically, and a specific mechanism was not proposed. Indeed, Planck proposes that interactions between oscillators and free electrons might be a possible cause for spontaneous emission, but this is not required for the statistical emission calculation. While this is a broad sense of reasoning about an open system, we can see here that Planck was implicitly treating the blackbody as an open system in continuous interaction with the environment. While this proposal was not ultimately retained as a consistent radiation theory, it is

interesting to note that it did give rise to the notion of zero-point energy, which became important in other contexts later on.

These examples of Planck's reasoning thus demonstrate how open and closed systems views can be fruitfully combined when modelling a system. However, they also seem to show that, in the context of newly developing rather than mature theories, scientists need not be specifically committed to one of the views. Nor is an exclusive commitment to one of the views necessary for making theoretical progress. In fact, this makes sense; Cuffaro and Hartmann's argument for the fundamentality of the open systems view is based on an analysis of highly developed theoretical frameworks, namely, alternatives for describing quantum systems. It is perhaps unsurprising that a more pragmatic, even opportunistic, attitude, should be taken in stages where a full theory has not yet been elaborated, as is the case in Planck's investigations of radiation. However, this provides evidence for the idea that one need not, indeed should not, be too committed to a single view in the context of emerging theories.

The non-exclusive use of the open and closed systems views can be seen as a wider example of using complementary reasoning strategies in a fruitful way. In the next section, we proceed to a consideration of the ensuing treatments of radiation theory by examining some of the historical reactions to it, namely by Planck's contemporaries, James Jeans and Albert Einstein. These reactions demonstrate another way in which complementary reasoning strategies are used in the context of a developing theory.

7.3 Reasoning about Quantisation

We saw in the previous section that Planck utilised both the open and closed systems views when developing his theory of blackbody radiation. Indeed, the combination was historically crucial, in that the two views correspond to the two assumptions which allowed him to develop his radiation theory. However, we often now designate these as being inconsistent assumptions. The obvious reason that we might be concerned about the argument given to arrive at the radiation formula is that in these assumptions, we have two statements that seem to contradict one another, namely, 'Resonators absorb radiation continuously', and 'Resonators absorb radiation in discrete integral quantities'. From the perspective of classical logic, this is enough to render any result unsound, insofar as we can derive anything from a contradiction. In fact, the literature does often describe this inconsistency in these terms. Consider, for instance, the following description.

> The inconsistency arises from the fact that in deriving [the radiation equation] theorists on the one hand appealed to a quantum postulate, according to which the energy levels of the 'resonators', or 'radiation oscillators' that give rise to the

> blackbody radiation, are quantized, but on the other hand they appealed to results drawn from classical electrodynamics, according to which these energy levels can vary continuously. (Saatsi 2014: 2951)

Similarly,

> In his work on black-body radiation, Planck combined a quantum hypothesis with classical electrodynamics, and came up with the first theoretical derivation of the empirical black-body law. Of course the quantum hypothesis was inconsistent with classical electrodynamics, and as Poincarré [sic] acerbically pointed out, we can derive anything we want from a contradiction. (Brown 2014: 3093)

While there was clearly something empirically correct about Planck's radiation formula, any broader theory of radiation based on such premises would seem to be, at the very least, problematic.[6] In what follows, though, I wish to draw attention to the fact that this explanation does not fully capture the reaction of either Jeans or Einstein. In this section, we will see in greater detail how these two physicists responded to the tension between the assumptions in their subsequent work, as well as how they proceeded to reason with these seemingly inconsistent assumptions.

In particular, I argue that a useful way of characterising the reasoning here is in terms of the identification of varying domains of applicability to which one can apply theoretical hypotheses. What is particularly notable in this case is that such a strategy can be understood in two complementary senses: on the one hand, we can narrow down, or restrict the domain of applicability of particular theoretical claims, in contrast with attempts to broaden the domain of applicability. The former is exemplified by Jeans's argument and corresponds to a practice that can be observed in numerous contexts, with Bohr's old quantum theory being a particularly relevant example. The latter can be observed in Einstein's approach, in what might be called a unifying or heuristic strategy when pursuing a new theory. The latter has also long been a prominent theme in philosophy of science. The notion of domains of applicability thus provides a way to understand both of these strategies, and so should be emphasised more strongly in describing scientific practice in the context of pursuing a new theory.

7.3.1 Jeans

Previous philosophical discussions of our case study, while insightful, have tended to focus on conceptual inconsistency rather than on more empirically motivated criticisms. As noted above, descriptions of Planck's work tend to identify the problem as the fact that Planck assigned the resonators conflicting properties,

or that the quantisation assumption conflicts with certain tenets of classical mechanics. Even the more historically-oriented analyses of Planck's thought sometimes present the problem in this way. For instance, Badino writes, 'Thus, the two ways of calculating the energy of the oscillator rested upon contradictory assumptions: classical continuous emission/absorption of energy in the first case, discrete elements in the second' (2012: A29). While it is true that Jeans noted this conceptual tension, I contend that it is valuable to consider the further argument he gives, which highlights the importance of empirical results, even in cases with a purported inconsistency.

Hudson (1989) and Gorham (1991) have described much of Jeans's work in the years following the introduction of Planck's radiation law, during which he tried to provide an essentially classical solution to the problem of blackbody radiation. By 1910 though, it seemed that he had come to the conclusion that this would be impossible, and turned his attention to defending the quantum theory. The first two chapters of Jeans's *Report on Radiation*, first published in 1914 (with a second edition issued in 1924), are committed to a demonstration of the fact that classical treatments of energy emission and absorption processes of a blackbody lead to the incorrect radiation law. In chapter 2, he considers the calculation of the radiation equation by dividing up the contributions of hypothetical processes contributing to the interaction of energy between matter and ether, including the vibrations of Planckian resonators, as well as interactions with free electrons. He considers each of these as separate mechanisms, and gives the expression that calculates their contribution to the energy of the system according to the idea that these are dynamical systems obeying classical laws (Jeans 1914: 14). For a system to have reached an equilibrium state, the partition of radiant energy for the system should be unaltered by interaction with an electron, so the equations expressing the partition of radiant energy for the system before and after its interaction should be equivalent. The resulting formula is not in agreement with experiment, indicating the need for a deviation from the classical formulas used to determine the contributions of these processes.

Of course, one way to deviate from these assumptions is by following Planck's reasoning. Nevertheless, Jeans found problematic the specific way that Planck justified his original principle of quantisation, which he expresses as 'The material energy of vibrating resonators can exist only in complete quanta' (Jeans 1924: 24). In the third chapter, he begins by explaining how one can calculate the probability that a system will have energy and momenta coordinates within some infinitesimal range around an amount ϵ in order to calculate the radiation formula for the system. Allowing the value of ϵ to approach zero expresses the idea that energy can take on any value and can be treated continuously. Evidently, this recovers the classical formula that is not in agreement with experiment, and the empirically observed formula is obtained by restricting the values of ϵ to be an integral multiple

of ν (Jeans 1924: 20–1). He then goes on to say that Planck's original method was based on different physical ideas, that of the Hertzian resonators.

Jeans outlines what would happen in a physical system corresponding to Planck's resonator model of the blackbody at equilibrium, where this body is interacting with ambient radiation (or ether). Specifically, he says, '[Planck's method] is open to serious objections. For, in considering the partition of energy between the various resonators, it is assumed that the energy can only vary by jumps of amount ϵ, while, in considering the partition of energy between resonators and ether, it has to be assumed … that the energy of the resonators can vary continuously' (Jeans 1924: 21). He adds that this assumption of continuous energy levels would then apply equally to the equilibrium condition for any interaction between ether and mechanisms such as free electrons, which we can expect to be present in many contexts. He goes on to explain that this is problematic because if it were the case, we would expect the resonators of the blackbody themselves to conform to Planck's radiation formula, but any mechanisms such as free electrons interacting with the radiant energy would tend towards the equipartition formula of classical mechanics. This is because the hypothesis of quantised energy was used to describe the energy levels of the resonators, whereas interactions between the ambient radiation bath and elements such as free electrons would be covered by the classical equations, just as Jeans described in the second chapter. Furthermore, note how he follows up on this analysis:

> The result would be a compromise between the two laws. Worse than this, it would be a compromise which would depend on the relative numbers of Planck resonators and of free electrons, and as the ratio of these would vary from one substance to another, there would be no definite law of radiation—the same for all substances—such as is demanded by observation. (Jeans 1924: 22)

That is, actual blackbodies may be composed of different materials, and ambient radiation conditions, including the presence of free electrons, will change from one setting to another. However, Kirchhoff's Law, which states that the radiation emission spectrum for blackbodies is independent of the constitution of those bodies, was an experimentally well-established principle. Thus, according to Jeans, the problem with these assumptions is not merely the conceptual incompatibility of assigning the resonators conflicting properties, but rather that doing so leads to a contradiction with such a principle.

Although Jeans seemed to take the somewhat standard view that Planck was attributing conflicting properties to the blackbody resonators, what is particularly interesting for our purposes is that a specific criticism of Planck's quantum hypothesis did not centre on the mere existence of inconsistent assumptions, but instead identified ways in which such contrary assumptions would result in incorrect

predictions. This would seem curious at first blush, if we were to take seriously the idea that a contradiction leads to any prediction whatsoever when considered strictly logically. Instead, I claim that we can recognise this as a case of reasoning by delimiting the domain of applicability of certain hypotheses in order to draw further conclusions. In this case, these conclusions serve to discredit the original hypothesis. The main observation that is relevant here is that this response is not strictly about the incompatibility of classical and quantum assumptions.

This criticism of Planck's derivation, while relatively minor in the scheme of Jeans's text, reveals several interesting points about pursuing a new theory. First of all, the fact that Jeans makes a specific prediction from contrary assumptions seems to give lie to the idea that from an inconsistent theory, one can derive absolutely anything. After all, as Vickers (2014) notes, 'just because one *can* derive anything and everything with deductive logic doesn't mean that there is a danger one *will*' (2014: 2900). To improve our understanding of the process of how we might reason about such a theory, though, it would be more useful to understand how these predictions were made.

What we see here is an implicit use of the identification of a domain of applicability. I suggest that, in the first place, Jeans's reasoning from these assumptions is much more similar to the kind of reasoning used by Bohr when he first introduced his atomic model for hydrogen, than to the more nebulous kind susceptible to the criticism of conceptual inconsistency. There is an enormous amount of literature on the inconsistency of the assumptions Bohr used in his first atomic theory, for similar reasons as with Planck's reasoning. However, he was clearly able to derive a prediction that could actually be compared with experimental results (1913). The 'inconsistency' is not one that went unnoticed in his case, yet it was possible to make certain predictions about spectral phenomena that were so precise that people took the results extremely seriously. Although Bohr used classical mechanics to describe the motion of electrons within a system, he also specifically posited the existence of stable states, and that those equations of motion were applicable only in these states. Thus, assumptions that might seem contradictory should actually be understood as applications of classical principles, that were well known to hold in certain domains, to a new but limited domain.

Similarly, in the case of Jeans's criticism, we can characterise his reasoning about the case as hypothesising about what would happen when limiting the applicability of the classical electromagnetic equations to the interactions between the radiation and free electrons in the blackbody, rather than allowing them to apply to all of the elements of the model. The difference between this and the Bohr case, of course, is that this particular restriction did not lead to results that accord with observation, and thus taken as an indication of a problem rather than an instance of confirmation. It is worth noting that this example demonstrates how such a technique is applicable in arguments for as well as against particular hypotheses.

7.3.2 Einstein

While Jeans focuses on the empirical implications for the system as a whole, Einstein highlights the significance of the quantum hypothesis for individual Planck resonators. As we will see, however, both are clearly examples that make specific inferences from the seemingly inconsistent situation to characterise the problem. More importantly still, Einstein's criticism calls into question whether there was indeed a genuine inconsistency. In this case, the reasoning on Einstein's part was arguably more fruitful, as it allowed him to pinpoint problems that actually had to be solved.

As explained in Section 7.2, the use of Maxwell's equations to calculate the mean energy of a resonator assumes that the energy of the resonators varies continuously. However, we should note that this is not equivalent to saying that the only way that a resonator could possess this mean energy value is to behave specifically in this continuous manner. In fact, it is consistent to posit that resonators behave in such a way that there are discrete jumps between energy levels, but that the mean energy nevertheless takes the value described by the continuous processes implicit in the classical picture. Planck himself seems to take into account such a possibility, when he says,

> It is true that this inconsistency is greatly reduced by the fact that, in reality, only mean values of energy are taken from classical electrodynamics, while, for the statistical calculation, the real values are used. (Planck 1913/1914: viii)

Of course, this raises the very important question of why discrete resonators would behave in such a way. Indeed, this seems to be exactly how Einstein addresses the flaw in Planck's reasoning. He starts by saying that, in Planck's theory, '[the underlying assumption is that] although Maxwell's theory is not applicable to elementary resonators, nevertheless the *mean* energy of an elementary resonator in a radiation space is equal to the energy calculated by means of Maxwell's theory of electricity' (Einstein 1906/1989: 196, emphasis in original). He then goes on to explain why this would be puzzling:

> This proposition would be immediately plausible if, in all those parts of the spectrum that are relevant for observation, $\epsilon = (R/N)\beta\nu$ were small compared with the mean energy $\bar{E}_\nu$ of a resonator; however, this is not at all the case ... It is easy to prove that according to Planck's theory of radiation, within the range of validity of Wien's radiation formula, $\bar{E}_\nu/\epsilon$ has the value $e^{-\beta\nu/T}$, thus, $\bar{E}_\nu$ is much smaller than ϵ. Therefore only a few resonators have energies different from zero. (Einstein 1906/1989: 196)

In other words, if the energy elements ϵ were small compared to the mean energy, it would be reasonable for calculations of the integrals to 'smooth out' so to

speak. However, in this case, the mean is based on individual values that are either zero, or quite large. Thus, the criticism here is that there is no reason to expect that the use of Maxwell's equations in such circumstances would yield the correct value for the average energy, and so it is in fact surprising that it does. Einstein seems to be drawing attention not to an impossibility, but to what might be seen as a strange coincidence that calls out for further investigation. In fact, Einstein does not even assign a conceptual incompatibility to the properties of resonators in this part of the argument, much less think that such an assignment of properties allows us to derive any result whatsoever. On the contrary, the criticism is intended to show that Planck was implicitly assuming the existence of something like light quanta. This claim then motivates the attempt to obtain the derivation in a way that does not rely on the seemingly conflicting assumptions.

While a useful way of understanding Jeans's criticism is in terms of narrowing down the domain of applicability of a hypothesis by restricting Planck's radiation law to a subset of the system being considered, I wish to draw attention to the fact that Einstein complements this strategy with a further heuristic approach. Namely, he considers instead how one might broaden the domain of applicability of the hypothesis of quantisation that led to the radiation law. In light of the foregoing considerations, Einstein assumes that there is something correct about Planck's quantum hypothesis and then, rather than restricting its applicability to merely resonators, he extends the domain of this hypothesis to contexts that were not yet considered in this light. One such application appears in his famous (1905/1967) paper on the creation and conversion of light, in which he suggests that the idea of quantisation should be applied to electromagnetic radiation itself rather than in the interaction of radiation with matter. He presented the major assumption of this paper as follows: 'According to the assumption considered here, when a light ray starting from a point is propagated, the energy is not continuously distributed over an ever increasing volume, but it consists of a finite number of energy quanta, localised in space, which move without being divided and which can be absorbed or emitted only as a whole' (p. 92). Then, in the latter part of his (1906/1989) paper on light, he uses the hypothesis of the existence of light quanta to derive a quantitative relationship between the Volta effect and photoelectric diffusion. The latter relationship was unrelated to light quanta up to this point, but Einstein's consideration of the effect of a quantum hypothesis in this context is an example of how broadening the domain of applicability leads to observable empirical results.

In fact, the former paper provides an example of how we can think in terms of both narrowing down and broadening the domain of applicability of a hypothesis. Einstein states that the wave theory of light has been 'excellently justified for the representation of purely optical phenomena' (p. 91), but notes that such phenomena deal largely with time averages rather than instantaneous values. Therefore, such a theory may very well lead to contradictions when 'applied to the phenomena of the creation and conversion of light' (p. 92). In other words, we ought to consider the possibility of narrowing the domain of applicability of

the classical wave theory of light to contexts that do not deal with its emission and absorption. Importantly, this reasoning strategy is then paired with the heuristic strategy to explore the results of applying a quantum hypothesis to various contexts. Importantly, this line of thought is extended to contexts beyond those specifically related to the emission and absorption of light: Einstein (1907) applies a quantum hypothesis to the context of explaining the specific heats of solids, thus significantly broadening the domain of applicability of the quantum hypothesis.[7]

7.4 Conclusion

In this chapter, I have considered certain aspects of Planck's seemingly inconsistent radiation theory, as well as ensuing criticisms of his theory, in order to articulate how an emerging theory can be well-served through 'complementary reasoning strategies' rather than strict rules. In particular, this historical example shows two different ways in which complementary approaches are fruitful. One is the combination of open and closed systems views, which I argue was operative in the context of Planck's articulation of his theory of radiation. The other is an example of reasoning by widening and narrowing the domain of applicability of a hypothesis under consideration. We saw in particular that a combination of open and closed systems views set the stage for the possibility of further criticism and development of the theory. In particular, the incorporation of an open systems view led to a fundamental change in the conception of the system, which in turn led to further theoretical developments. Of course, there are significant differences in the two examples. In the first, two views are combined into one theory. In the second, we see how complementary strategies can lead to fruitful discussions and criticisms when scientists engage with a newly proposed theory. Nevertheless, I would argue that both are important. In particular, in our case study, where available information was limited and a coherent theory had not yet been established, the combination of these strategies was crucial for pushing inquiry forward.

Endnotes

1. For instance, see Meheus & Nickles (2009) and Schickore & Steinle (2006) for literature on scientific 'discovery', or Achinstein (1993) for an example of 'theory pursuit'.
2. One way to physically instantiate this situation is by considering an enclosure with a small hole through which electromagnetic radiation enters. Upon entering the enclosure, the radiation is reflected among the walls of the body, creating a radiation bath that can be brought into thermal equilibrium.
3. See Darrigol (1992) for a history of the development of the blackbody.
4. Unfortunately, it implied that the radiation emitted by a blackbody would be infinite at certain temperatures. This problem came to be known as the 'ultraviolet catastrophe', and was taken to indicate serious problems for classical physical theories. However, Klein (1961) argues that it is unlikely that this more general problem played a major role in Planck's original reasoning.
5. See Badino (2012) for further discussion of this theory.

6. To be clear, the above-cited authors do not subsequently subscribe to the idea that Planck's theory is suspect simply because of the fact that anything might follow from such a contradiction, as the above observations are made in the context of their own philosophical analyses, Saatsi in an argument for scientific realism, and Brown on the nature of evidence. I cite these passages merely to underline the characterisation of the inconsistency that is often given in the literature.
7. It is worth mentioning that another proponent of the latter strategy can be found in Henri Poincaré, who first encountered Planck's quantum theory at the first Solvay Congress (as detailed in McCormmach 1967) and later calls such a conception 'fascinating' due to the fact that there was a trend towards atomic theory in multiple domains of physics (Poincaré 1913/1963: 83).

Bibliography

Achinstein, P. (1993), 'How to Defend a Theory Without Testing It: Niels Bohr and the "Logic of Pursuit"', in *Midwest Studies in Philosophy* XVIII: 90–120.

Badino, M. (2012), 'Productive Conservatism: Planck's Second Theory of Radiation', in *Annalen der Physik* 574: A29–A31.

Badino, M. (2015), *The Bumpy Road: Max Planck from Radiation Theory to the Quantum (1896–1906)* (Springer).

Bohr, N. (1913), 'On the Constitution of Atoms and Molecules, Part I', in *Philosophical Magazine* 26: 1–25.

Brown, M. B. (2014), 'The Shape of Science', in *Synthese* 191: 3079–109.

Cuffaro, M. E., & Hartmann, S. (2024), 'The Open Systems View', in *Philosophy of Physics* 2: 1–27.

Darrigol, O. (1992), *From c-Numbers to q-Numbers: The Classical Analogy in the History of Quantum Theory* (Berkeley: University of California Press).

Einstein, A. (1905/1967), 'On a Heuristic Point of View about the Creation and Conversion of Light', trans. D. ter Haar, in *The Old Quantum Theory: Selected Readings in Physics* (Great Britain: Pergamon Press). Original work published 1905.

Einstein, A. (1906/1989), 'On the Theory of Light Production and Light Absorption', trans. A. Beck, in *The Collected Papers of Albert Einstein, Volume 2: The Swiss Years: Writings, 1900–1909 (English translation supplement)* (Princeton University Press), 192–9. Original work published 1906.

Einstein, A. (1907), 'Die Plancksche Theorie der Strahlung und die Theorie der spezifischen Wärme', in *Annalen der Physik* 22: 180–90.

Gorham, G. (1991), 'Planck's Principle and Jeans's Conversion', in *Studies in History and Philosophy of Science* 22: 471–97.

Hudson, R. (1989), 'James Jeans and Radiation Theory', in *Studies in History and Philosophy of Science Part A* 20: 57–76.

Jeans, J. H. (1924), *Report on Radiation and the Quantum-Theory* (London: Fleetway Press, Ltd.). Second edition.

Kirchhoff, G. (1860), 'On the Relation Between the Radiating and Absorbing Powers of Different Bodies for Light and Heat', in *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 20. Translated by F. Guthrie from *Annalen der Physik* 109: 275–301.

Klein, M. J. (1961), 'Max Planck and the Beginnings of the Quantum Theory', *Archive for History of Exact Sciences* 1: 459–79.

Lummer, O., and Pringsheim, E. (1900), 'Über die Strahlung des schwarzen Körpers für lange Wellen', in *Verhandlungen der Deutschen Physikalische Gesellschaft* 2: 163–80.

McCormmach, R. (1967), 'Henri Poincaré and the Quantum Theory', in *Isis* 58: 37–55.

Meheus, J., & Nickles, T. (eds.) (2009), *Models of Discovery and Creativity* (Springer).
Planck, M. (1900/1972), 'On an Improvement of Wien's Equation for the Spectrum', trans. D. ter Haar, in *The Old Quantum Theory* (Oxford: Pergamon Press), 79–81.
Planck, M. (1912), 'Über die Begründung des Gesetzes der schwarzen Strahlung', in *Annalen der Physik* 342: 642–56.
Planck, M. (1913/1914), *The Theory of Heat Radiation*, trans. M. Masius (Philadelphia, PA: P. Blakiston's Son & Co.). Second edition.
Poincaré, H. (1913/1963), *Mathematics and Science: Last Essays (Dernières pensées)*, trans. J. W. Bolduc (New York: Dover Publications, Inc.).
Saatsi, J. (2014), 'Inconsistency and Scientific Realism', in *Synthese* 191: 2941–55.
Schickore, J., & Steinle, F. (eds.) (2006), *Revisiting Discovery and Justification: Historical and Philosophical Perspectives on the Context Distinction* (Dordrecht: Springer).
Vickers, P. (2014), 'Theory Flexibility and Inconsistency in Science', in *Synthese* 191: 2891–906.

8

Boundaries, Frames, and the Issue of Physical Covariance

Henrique Gomes, Simon Langenscheidt, and Daniele Oriti

8.1 Introduction

In this contribution, we focus on two kinds of idealisations that are routinely employed when modelling physical systems in a gravitational context. Namely, that of negligible spacetime boundaries (often associated with the restriction to closed systems) and that of ideal reference frames. In this chapter, we will investigate what happens upon the removal of these idealisations, i.e., when we make our models more realistic.

In doing so, we think very carefully about the modelling process itself, and specifically take seriously recent developments in classical and quantum-gravitational physics as well as quantum foundations, pointing towards some of their nontrivial implications. These have to do with gravitational (more generally, gauge) systems in finite spacetime regions, relational dynamics in classical and quantum gravity, and quantum reference frames. Lastly, we will tie together the main lessons we gather from such developments and conclude that, when the above idealisations are removed, we are left with a description of gravitational systems which is necessarily perspectival through a dependence on the adopted physical reference frame. This, in turn, is understood as the (partial) embodiment of the agent modelling the system or (depending on the context) the observer accessing it to determine its spatiotemporal properties.[1] We also emphasise that this 'perspectivalism' has the consequence that relating the descriptions made by different observers is no longer a straightforward matter.

Before we delve into our main arguments and the conclusion they lead us to, we introduce a bit of context and broadly motivate our reasoning and our attention to the issue of boundaries and frames in physics, and gravitational theory in particular. We are sympathetic to operational attitudes towards physical modelling and believe that physical models should, as much as possible, reflect the realistic conditions of our interactions with the relevant physical systems. We consider this an important component, albeit not the only relevant one, of scientific methodology. We understand the modelling of the system to be an activity that is performed by

Henrique Gomes, Simon Langenscheidt, and Daniele Oriti, *Boundaries, Frames, and the Issue of Physical Covariance*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press. © Oxford University Press (2026). DOI: 10.1093/9780198929277.003.0009

a variety of types of agents (Barzegar et al. 2023). For this activity to make sense, there must be an initial separation between the agent, including any equipment relevant to the modelling activity (recording devices, etc.), and the system that is being modelled by the agent. Such a split must also be reasonably stable under the dynamics of both agent and modelled system.

Particularly, among their equipment, we focus on anything used to define a reference frame for spatiotemporal properties: other physical systems that the agent uses to project properties of the modelled system into a form that could encode its spacetime localisation, e.g., a set of chosen clocks and rods. These reference frames serve, of course, also the purpose of contextualising the agent itself within the larger system it constitutes together with the modelled system. It is through their reference frame that an agent acquires a perspective on the full system and its properties. One issue we will discuss is whether the reference frame itself affects said perspective in ways that are ultimately non-negligible.

A central focus of our discussion is how physical systems defining reference frames are actually modelled for a specific gravitational system (which could be the spacetime geometry itself), and how many of their own physical properties are accounted for in the model. Indeed, one can (and, we will argue, should) include such a reference frame explicitly within the model of the system of interest. The issue then becomes how this inclusion, when performed in less idealised terms, changes the description of the modelled system, and what the conceptual implications of these changes are.

A second, related way in which the adoption of an agent/observer-based perspective influences our modelling of gravitational systems has to do with the resulting localisation of the modelled system in spacetime, regardless (initially) of how this localisation is determined via reference frames. The modelled system can be seen as occupying a specific portion of spacetime, and this portion does not include the observer/agent herself, i.e., it is not the same portion of spacetime that the observer/agent occupies (as she does not appear in the model of the system). The observer–system split, therefore, is also reflected in the existence of a boundary of spacetime, separating the two. These boundaries typically take the form of some timelike or lightlike submanifolds which enclose either the observer or the system (up to being possibly 'capped off' by some spacelike initial or final surfaces).

Neglecting the role of agents/observers and the existence of such a split translates, in our models, into the treatment of the system as closed and also in its description as occupying a dynamically closed region of spacetime, modelled as a closed manifold without boundary, or as a manifold with boundary but with special boundary conditions preventing information or degrees of freedom from leaving or entering the system. The inclusion of spacetime boundaries in our models, moreover, can be done in more realistic or more idealised terms, and the ensuing question is, again, how removing idealisations affects the act of modelling itself.

So what happens when, in our models, we do not neglect the presence of spacetime boundaries? As we will argue, their presence is relevant as they affect the physics of the systems of interest; and this is closely tied to the issue of modelling, in a less idealised way, physical reference frames. In the case of spacetime boundaries, the idealisations involved are of two types. The more drastic one is to simply neglect the existence of boundaries and treat the system of interest as closed even in its spacetime and gravitational properties, as we discussed. Less drastic, but still an idealisation, is the assumption that the spacetime boundary (thus, the observer) can be 'pushed to infinity', i.e., the imposition of asymptotic boundary conditions and of infinite distance/volume limits. In a gravitational context, this is often accompanied by the switching off of the gravitational interaction between the system of interest and the system 'outside the boundary', when one imposes flat or weak gravity data at the asymptotic boundary. The growing interest in gravitational systems in finite regions has resulted in a large number of recent developments, which in turn have improved also our understanding of the asymptotic setting, with both foundational and possibly observational consequences. Moreover, the physics of gravitational systems in finite regions ties closely to the physics of reference frames and relational dynamics, as we will discuss in the following.

We therefore arrive at our main point. A more realistic understanding of coupling of (gravitational) systems, going beyond typical idealisations, and the construction of physical questions, as represented by gauge-invariant observables, points to the non-negligible role of physical frames (as the partial embodiments of observers) in the understanding of gravitational systems, at classical and even more so at quantum levels. Furthermore, this has implications for what we should expect in terms of comparisons and translations of statements made by different observers, and it suggests that such physical understanding is of an irreducibly perspectival nature.

We will now present in more detail several arguments for these ideas, and give pointers to relevant ongoing research directions, before discussing briefly some of the broader conclusions we draw from them.

8.2 The Issue of Physical (Non)Covariance

Drawing on a solid body of results from mathematical gravitational physics, quantum gravity and foundations of physics, the main conclusion we want to argue for, in this contribution, is the following. Once we remove two standard idealisations from our models of gravitational systems, whether classical or quantum, and in particular, once we take seriously the distinction between open and closed systems (encoded in the most realistic way in the presence of finite spacetime boundaries), as well as the need for physical reference frames (encoded in gauge invariance

and nonlocality with respect to the supporting manifold), we realise that a crucial issue remains open in fundamental gravitational physics: that of physical covariance with respect to transformations across physical, realistic reference frames, together with the question of what, exactly, is in fact invariant under the same transformations. Our reasoning points to the inevitability of this issue, and thus calls for devoting to it our full research attention.

Before presenting the main steps in our argument, and their supporting evidence, we state the basic presumptions we are starting from.

(I) Modelling conventionally starts with a system–agent or system–observer split. Modelling a physical system requires one to first identify all the data and observables referring to the system and then to set up an epistemic boundary between the system itself and the agent performing the modelling. By such a boundary, we mean both a conceptual or abstract split of all of the degrees of freedom into independent parts that are, to a certain extent, controllable by the agent, as well as any concrete physical object or phenomenon that implements such a split on the level of the actual system. This provides some separation of the agent from the system to be modelled, i.e., for all properties whose exclusion does not impair the explanatory or predictive accuracy of the model.

Such properties may, however, as we will argue, possibly include some properties of the agents themselves. Therefore, we presume that the split is at first more a matter of convenience, and that the agent can, within reason, include their own properties and dynamics in the model of the system, when these are considered relevant. If the system is to be also modelled in its spatiotemporal properties—and this is clearly necessarily the case for gravitational systems and, in fact, extends to any dynamical system—its spacetime localisation is encoded by assigning it a determined portion of the manifold we conventionally use to represent spacetime. The existence of a system–observer split is then encoded in the presence of a boundary for such a region, with this boundary being finite in realistic models (and before any further approximation). This is the case both if the model represents the observer as occupying a finite spacetime region, with the system constituting whatever is 'left over', or, as is more customary, if it is the system that is being modelled as occupying a finite spacetime region (the observer being described as part of 'the environment').

(II) We also assume that modelling activity may involve a general type of reference frame. By a reference frame, we generally mean any, possibly entirely conceptual, entities used as a way by which an agent standardises their recordings of phenomena in the system. The simplest example of this, in the context of spacetime physics and gravitational systems, is

a coordinate frame, which consists of a set of conventional real labels the agent assigns (in principle arbitrarily) to the points in the manifold mentioned before, interpreted as an encoding of the temporal and spatial localisation of the 'event associated with that point', for instance, a model of the reading of some clocks and rods. By fixing such a frame, the agent can organise their recordings by projecting them into a set of numbers respective to the frame, for example the mentioned coordinates. We can speak further of a reference frame as *physical*, which we understand as any set of agent-controllable or -accessible degrees of freedom which can serve as a means to situate the agent within the full system, and serve as a standard to record measurements about the system (thus, they are frames according to the general definition), but are modelled by accounting explicitly for the dynamical and physical properties of the actual subsystems used by the agent to measure/access (at least indirectly or hypothetically) the system. In the case of clocks and rods, of interest for what follows, they translate in the model into a physical reference frame to the extent that their actual physical properties as matter systems are accounted for.

(III) It is possible, interesting, and often necessary, to include aspects of observers/agents in a model of a system of interest. One could say that the abstract observer/agent has been 'embodied' in specific physical systems or mathematical conditions (e.g., boundary conditions, external potentials, etc.) appearing in the model. In particular, an aspect of an observer/agent that can be embodied in the model corresponds exactly to the spatiotemporal reference frames, i.e., the set of clocks and rods used to localise the physical system of interest in space and time (in fact, one could say that these devices *define* such a localisation). This is where the issue of how realistically reference frames are modelled becomes relevant, and with it, in particular, the distinction between coordinate and physical reference frames. This distinction will be a key point in our argument below. To anticipate aspects of our argument, if reference frames are to be considered to be a part of physical reality, and not just abstract elements in the modelling process, their physical properties, including their interaction with other physical systems and their backreaction on spacetime geometry, should be carefully modelled and quantified (in particular if they are to be modelled within a theory of quantum gravity and quantum spacetime). At the quantum level, the quantum properties of the physical systems used as reference frames should also be accounted for in the model.

Now, on the basis of these presumptions, let us present our arguments in a way that builds up to the conclusion we want to emphasise.

0. Starting point

The localisation of physical systems in spacetime and their dynamical evolution through it are conventionally modelled in terms of points and trajectories in a differentiable manifold, and made explicit, in each region of such a manifold, by their expression in terms of coordinate frames. Their domain of definition identifies the spacetime regions which can constitute the domain of the physical systems under consideration, often assumed to run up to infinity in time and/or space. Both the association of coordinate frames with observers (as they are defined/assigned by them), or, better, the modelling of the reference frames used by the observers as coordinate frames, and their extending to infinity encode two important idealisations, on whose removal from our models we focus in the following.

Note that steps 1 and 2 will contain most of the details required to construct the remaining arguments, and are for that reason longer and more technical.

1. Diffeomorphism invariance implies that coordinate frames are unphysical

Conceptually, the gauge symmetries of gravitational theories, specifically diffeomorphism invariance, leads to the view that coordinate frames are entirely unphysical: they are elements of the formalism (of manifolds and modelling) that do not affect physical results and that have no physical significance. Of course, the use of coordinates has enormous practical advantages, but it should be recognised that this does not develop a foundational role for them in our understanding of gravitational systems (and by extension, of the physical world). In fact, full invariance under active diffeomorphisms runs deeper than just coordinate invariance (Gaul & Rovelli 2000), and it has important consequences for the dynamical aspects of spacetime in General Relativity(GR)–like theories (Giulini 2007).

Whether this lack of physical status for coordinate reference frames extends to points and trajectories in the manifold, and thus to the manifold itself, is more subtle. One could hold the view that what is physical must be gauge-invariant. In this light, one concludes that bare manifold points (thus directions, subregions, etc.) carry no physical significance, since there are diffeomorphisms that are symmetries of the theory that won't preserve the given subregions. However, even adopting this view, one may still wonder about the status of quantities entering into the definition of gauge-invariant observables; for instance, field values whose relations to other field values may together constitute a gauge-invariant observable (Rovelli 2002). Since no function that carries manifold points as free (or unbound) variables is, in itself, a gauge-invariant observable (diffeomorphisms map points to other points and thus nontrivially transform such functions), and this includes any set of field values, such parts of an invariant observable need not be themselves invariant: they only need to have the right covariance property. Thus all invariant observables need to be instantiated by suitable diffeomorphism-invariant functions of field values (Dittrich 2006, Goeller et al. 2022), functions in which the

bare manifold points either don't appear or figure only as bound variables (in this limited sense, reference to manifold points can be considered unproblematic). Such functions could be the relational observables we discuss below.

In any case, it is conceivable that manifold points, directions, paths, etc., may altogether disappear from the formulation of our models, and may not feature at all in more fundamental theories, i.e., in quantum gravity. Explicit examples of such a disappearance are given in classical GR, when this is deparametrised in terms of special material frames (Brown & Kuchar 1995; Giesel & Thiemann 2010), and in a handful of approaches to quantum gravity, in various approximations (Marchetti & Oriti 2022; Giesel & Thiemann 2015), and in toy models like those that appear in quantum cosmology (Banerjee et al. 2012).

Technically, we can associate these issues regarding observables to the mathematical formulation of gauge theories as follows. In gauge theories and in gravity, dynamical symmetries of the basic variables involve spatiotemporal derivatives of arbitrary functions. It follows that the symmetry-invariant content of the initial state can be extracted only nonlocally from the basic variables, from procedures that involve spacetime points only as bound variables, e.g., as variables being integrated over. In more detail, a local symmetry implies, via Noether's second theorem (see Brading & Brown 2000 for a conceptual exposition), that the equations of motion of the theory are not all independent: the full set of Euler–Lagrange equations is not linearly independent, or, alternatively, it obeys constraints. Thus this set of equations only *uniquely* determines the evolution of a subset of the original degrees of freedom. This subset can be interpreted as made of 'composite' variables: it is described by relations between the original degrees of freedom. The evolution of the remaining degrees of freedom is arbitrary. These remaining degrees of freedom are usually taken to not describe any physical feature of the system and are called *gauge*. There is thus a certain freedom in choosing which composites of, or relations between, the original fields will be uniquely propagated, or will evolve deterministically—each choice corresponds to *a choice of gauge*.

In the Hamiltonian phase space formalism, this indeterminism comes in the form of first-class constraints that physical observables have to satisfy. A choice of physical degrees of freedom is a choice of canonical degrees of freedom that satisfy the constraints, which amount to choices of phase space functions that completely fix the gauge freedom (and, in Hamiltonian analysis, are phase space functions that form a 'second-class' system together with the constraints). In the next item, we will see how these technical features are associated to a (benign) form of nonlocality.

2. Localisation in space and relational time requires physical frames

The implication of gauge symmetries for observables aligns, in fact, with a more realistic (and operational) description of spacetime localisation and evolution. Observers use physical (material) clocks and rods to define them, so physical

observables are best understood as relations between dynamical fields and the (material) clocks and rods used for their localisation in spacetime. This is the relational strategy for the definition of physical, thus diffeomorphism-invariant, observables in gravitational physics (Rovelli 2002; Dittrich 2006). Their construction is often difficult, but their in principle availability is a fact. It is a fact, that is, provided one seeks them to be defined on subregions of the supporting differentiable manifold only, and not globally, and accepts the need to use several physical frames for a global description of physical systems throughout space and time. State-of-the-art explicit construction for this can be found in Goeller et al. (2022). Explicit examples of such relational constructions are given in classical GR, deparametrised in terms of special physical frames (Brown & Kuchar 1995; Giesel & Thiemann 2010), and in a handful of approaches to quantum gravity, in various approximations (Marchetti & Oriti 2022; Giesel & Thiemann 2015), and especially in toy models like those that appear in quantum cosmology (Banerjee et al. 2012; Hoehn & Vanrietvelde 2020).

One way to find such descriptions in practice, what we alluded to above as 'choosing a gauge', is also known as *gauge fixing*. Gauge-fixing functions can often be interpreted as characterising a *reference frame* for the gauge group in question. For example, in Newtonian mechanics, a gauge-fixing of translations is given by fixing the centre of mass to be immobile, and this fixes the frame to be the centre-of-mass frame, which is a thoroughly relational entity. Note that the Newtonian example does not illustrate the main feature that we are describing here: indeterminism of the equations of motion. The reason is that classical non-relativistic particle mechanics does not admit time-dependent symmetries and so does not suffer from indeterminism (see, e.g., Wallace 2002). Nonetheless, the example is similar in the relevant aspects to gauge-fixing in general relativity or Yang–Mills theory.

Technically, the composite or relational quantities that emerge from gauge-fixing are usually nonlocal (with respect to the supporting manifold) because, in general relativity and Yang–Mills theory, choices of initial data must satisfy elliptic equations, whose solutions require integrals over an initial spacelike surface. In practice, ellipticity means that boundary value problems require only the boundary configuration of the field, i.e., they do not also require the field's rate of change at the boundary. Thus the solution of these equations exists *on each simultaneity surface*, and so the solution does not describe the propagation of a field, as would a solution to a hyperbolic equation. Thus nonlocality may arise because the function that takes the original local degrees of freedom to a uniquely propagated subset is, generally, nonlocal: the value of an element in the subset at point x depends on the values of the original degrees of freedom at other points.

But this nonlocality can also often be understood as emerging from relations to a reference frame: for instance, geodesic distance relative to a boundary in a geodesically convex region would instantiate nonlocality, and would also fix the

spatial diffeomorphisms of the initial surface relative to the fixed boundary (cf. Gomes & Riello 2017; Carrozza & Höhn 2022; Carrozza et al. 2024). Once we describe all fields relative to such a reference frame, they are better seen as 'composite' or 'dressed' by functions of all of the fields: they can be then expressed as gauge-invariant functions.

But we can bypass the mention of reference frames and obtain a straightforward equivalence between gauge-fixings and invariant, composite, or 'dressed' functions by following Gomes (2024b). We quickly summarise the method. Given a gauge-fixing function C, there is a gauge transformation γ that takes any given configuration φ into one satisfying the gauge fixing, $C = 0$. The mapping $\varphi \mapsto \gamma(\varphi)$ is then referred to as *the dressing*, and in the case of gauge-fixing functions that involve derivatives it usually involves integrals and is therefore nonlocal. One can then verify that the *dressed* fields, $\varphi^{\gamma(\varphi)}$ are gauge-invariant quantities.

Now, one can also understand the dressing $\gamma(\varphi)$ as implicitly tied to a reference frame as follows: it is the transformation—e.g., the diffeomorphism—that takes an arbitrary representation of the state to the representation according to the reference frame. To illustrate, suppose we have four scalar fields that are linearly independent in some region; say, four solutions of Klein–Gordon equations for different initial values, call these $\phi^{(I)}$, where $I = 1, \ldots, 4$. Thus, putting the four together into a 4-tuple, we have a local diffeomorphism, for $U \subset M$, $\bar{U} \subset \mathbb{R}^4$, $\phi^{(I)} : U \to \bar{U}$. Of course, since these are physical fields, they will appropriately covary under the action of the diffeomorphisms: this is the property required of γ above. Now, using this map, we can write: $g_{IJ}(\phi) := \left[(\phi^{(I)})^{-1}\right]^* g_{ab}$, where with $\left[\,\cdot\,\right]^*$ we denote the pullback and with $g_{IJ}(\phi)$ the components in the frame $\{\phi^{(I)}\}$ of the abstract metric tensor g_{ab}. Such a quantity is commonly defined a *relational observable* and it is clearly invariant under the active diffeomorphisms. (See Goeller et al. 2022; Kabel et al. 2024; Bamonti 2023; and Bamonti & Gomes 2024 for more about such examples and their relation to gauge-fixings and dressings.)

In addition to these facts about the dynamics of gauge theories, we can give two further points about the construction of invariant observables. (i) Both gravitational and material degrees of freedom can be employed in the construction of reference frames/dressings/observables. (ii) Reference frames, or a choice of relational observables, etc., will in both general relativity and Yang–Mills theory only be available *locally* in field space. In the gauge-fixing language, this is a consequence of the *Gribov obstruction* (Gribov 1978; Singer 1978). Conceptually, the obstruction emerges from the physical nature of the reference frame: for some field values, the configuration that made up the reference frame simply won't have sufficient structure to serve its role as a reference frame.

Let us add a couple of remarks. The nonlocality we have been referring to in the above discussion is with respect to the manifold supporting the fields. This nonlocality could be seen as a further suggestion that manifold points have indeed no direct physical significance, albeit they may enter indirectly in the construction

of physical (diffeomorphism-invariant) quantities. Indeed, in the relational strategy, physically meaningful spacetime localisation is always with respect to field values, i.e., those fields used as reference frames, and local physics is actually to be defined with respect to such field values (Goeller et al. 2022). One could say that local physics takes place in field space, not the supporting manifold. This is one motivation to consider dynamics in such field spaces as the relevant arena for extracting physical effects from quantum gravity models (Oriti 2024).

3. Physical frames are (very) different from coordinate frames

The conclusion of the relational strategy for the construction of diffeomorphism-invariant observables and a physical definition of spacetime localisation is that it requires the use of physical reference frames, constituted by dynamical entities (fields) in the model. Obviously, this is not only in line with the formal requirements of the gauge invariance, but also with the lesson of background independence.

It is important to realise that this step brings truly novel features into our models of gravitational systems, and not just formal adequacy with the mathematical requirements. Physical reference frames possess properties that coordinate frames do not possess, and such properties have to be carefully accounted for in our models, otherwise we will pay the price of missing out on a number of potentially crucial physical effects.

Let us illustrate these properties by stressing the difference with coordinate frames. Physical frames, in general, can be adopted only 'locally', i.e., for a limited range of values of the fields that we interpret as representing temporal and spatial distances, just like coordinate charts in general only make sense locally on the manifold. As with standard coordinate charts, this is markedly the case on manifolds with nontrivial topology. But, as we mentioned above when discussing the Gribov ambiguity, unlike coordinate charts, the limitations of applicability of physical reference frames can also be due to dynamical considerations, e.g., a certain set of fields can be only used as a reference frame for a subset of solutions of its equations of motion, or for a restricted region in field space. More crucially, since physical reference frames are themselves fully dynamical, they must interact with other dynamical entities in the model. In particular, they are coupled to geometry via Einstein's equations (or those of other diffeomorphism-invariant gravitational theories), thus they backreact on the geometry in a way that depends on their own energy–momentum tensor. This backreaction may not be negligible. In general, these and other physical effects affect the gravitational dynamics of spacetime and matter described in terms of physical reference frames (Giesel et al. 2021).

The above is the most immediate (interesting) complication arising from using physical frames, in the classical domain. Being truly physical, however, dynamical reference frames are also quantum systems. This means that, in a truly realistic model of our gravitational systems, accounting for diffeo-invariance via

the relational strategy, our clocks and rods should be described as any other physical system in the model, with their own Hilbert space of states, their algebra of observables etc. It is intuitively clear that this brings a whole new set of complications, which are not merely formal, and which are being carefully addressed in recent research (see Kabel et al. 2024; de la Hamette & Galley 2020; Hoehn et al. 2021b; and references therein). To list a few: quantum reference frames will generically be in a superposition of the states corresponding to definite values of the observable we identified as our clock or rod reading, thus such readings will be subject to quantum fluctuations; quantum reference frames will be entangled with other physical systems; quantum reference frames will be subject to contextuality conditions; typical notions like the relative state of the system being in a superposition or carrying entanglement between subsystems depend crucially on which reference frame's point of view is used.

In particular, as soon as one considers *multiple* observables relating to a given frame, one cannot in general make all of them sharp in a given state (as opposed to a single observable, where this is always possible). The above is generic, but it has immediate implications for spacetime localisation. For example, we may expect that incompatibility conditions will affect those observables that we use to define spatiotemporal localisation (generically requiring multiple observables), resulting in some form of spacetime non-commutativity. Generally speaking, the proper use of quantum reference frames in gravitational physics has been proven to inform and clarify our notions of localiseability (Giacomini 2021; Chataignier et al. 2024), time evolution (Hoehn et al. 2023; Hausmann et al. 2024), gravitational entropy (Vuyst et al. 2024) and the wave function of the universe (Hoehn 2019). Moreover, quantum uncertainty could make it unreasonable to neglect the backreaction of physical frames on the geometry.

Take the simple example of a scalar field ϕ used as a clock, in a Hamiltonian framework. Its conjugate variable, the momentum π_ϕ, can be understood as a contributor to the energy density of the field.[2] Therefore, at a classical level, one must keep said momentum as small as possible, in order to allow for a suitable evolution of the value of the clock field while keeping gravitational backreaction to a minimum. However, if the field is quantum, this is no longer an option. Due to the uncertainty principle, there is complementarity between the two variables and one cannot specify them both in *any* state to an arbitrary degree of sharpness. Particularly, naively making the determination of the clock value sharper and sharper induces large uncertainties in the value of the momentum. This means that it is no longer accurate to neglect the backreaction caused by the field (because a small expectation value of the clock value is overwhelmed by fluctuations, which involve possibly quite large energy densities participating in the quantum interaction of the clock with the gravitational field, and which could induce similarly large fluctuations in geometry too).

All the above has to be taken into account in models of quantum gravity where gravitational dynamics is studied (and extracted) in relation to physical frames (Marchetti & Oriti 2022).

4. Coordinate frames are physical only in (useful) idealisations, which have to be ultimately removed

Coordinate frames are the result of an idealisation (which is technically implemented by various approximations). They correspond to specific reference frames used by agents/observers, and can be obtained as a limiting case of physical reference frames, modelling them (using the terminology explained above), and thus physical degrees of freedom, in the idealised case in which they are modelled as not gravitating, not interacting with other physical systems, and classical (Dittrich 2006; Giesel & Thiemann 2015). Removing this idealisation requires one to further develop the construction and analysis of physical (diffeomorphism-invariant, relational) observables in classical and quantum gravity, and the construction and analysis of quantum reference frames. It also requires gaining proper control over the coupled dynamics of such physical reference frames and other systems, including spacetime itself.

The success of the routine use of coordinate frames in physics (including gravitational physics) shows that the approximations leading to idealised frames are applicable and robust in most regimes of interest. The point about them being idealised remains, however, and we should consider carefully and critically such approximations in the context of models in which they are not assumed from the start. We could be missing interesting new physics. In particular, in regimes in which quantum gravity is relevant, we should not expect such approximations to be valid, in general.

There are two cases, however, in which standard coordinate frames can be physical and diffeomorphism invariant, i.e., in which it is not apparent from the mathematical structure of the model that these frames are the result of a drastic idealisation from some more appropriate modelling of the physical reference frames (used by the agent/observer). In the first, the spacetime geometries being considered possess special isometries. They thus carry a notion of preferred observers and spacetime directions. In the second case, the gravitational system is modelled in terms of spacetime with boundaries at infinity and corresponding special asymptotic boundary conditions. While these situations are indeed special cases, corresponding to the idealised case that neglects entirely the quantum features of the system, and therefore do not alter the conceptual significance of the above conclusion, which applies to the general case, it is important to realise that another crucial idealisation is in place in the second type of situation as well.

Indeed, to assume that the gravitational system under consideration occupies an infinite region of spacetime, and that the observer studying it can sit at an

asymptotically safe distance from it, is, clearly, another (useful) idealisation. It corresponds to the case in which we consider a system that is actually localised in some finite spacetime region that just happens to be big enough relative to the physical reference frame to be idealised as infinite. Removing this idealisation and thus making our models of gravitational systems more realistic, means tackling a number of challenges and further developing a research direction that has recently gained much attention: that of gravitational physics in finite regions, and the novel features that are brought into relevance by the presence of finite boundaries. A general result is that such boundaries should be endowed with dynamical edge modes, also in order to preserve full gauge invariance, when diffeomorphisms are taken to act on boundary degrees of freedom. A second result is that edge modes of such regions can be seen as providing a physical reference frame, so that defining a physical spacetime boundary in all its dynamical aspects is equivalent to selecting a physical frame. These two results form the basis for the next two steps in our argument.

5. Considering gravitational physics in finite regions brings in further physical features

Gauge symmetries in a bounded subsystem are quite subtle. Dynamical structures for the subsystem, such as its intrinsic Hamiltonian, symplectic structure, and variational principles in general, acquire terms that must be carefully dealt with.

The conceptual reason for the subtle treatment of gauge symmetry in subsystems is that gauge theories manifest a type of nonlocality, as we described above. Thus the global, physical phase space (or the corresponding global physical Hilbert space, in the quantum theory) is not factorisable into the physical phase spaces over the composing regions.[3]

The technical reason is that, setting aside very stringent physical boundary conditions—such as vanishing field-strength in the case of Yang–Mills theory—the variation of the action functional produces boundary terms that are not gauge-invariant. This is easy to see: any Lagrangian that employs variables that are not gauge-invariant—and this includes both general relativity and gauge theory—will, upon variation, produce boundary terms that depend on the variation of these variables. On the initial value surface, such terms are innocuous, and will merely inform us of the symplectic structure of the theory: which variables are canonically conjugate. But for timelike boundaries, we obtain non-invariant terms. How should we interpret these?

We can easily illustrate this with the Yang–Mills action in vacuum (capital Roman letters refer to Lie algebra indices):

$$S(A) := \int_{M\times\mathbb{R}} F^{I}_{\mu\nu} F_{I}^{\mu\nu}. \tag{1}$$

On a bounded submanifold, say, $R \times \mathbb{R}$, where $R \subset M$ is a spatial submanifold of M and $S = \partial R$, a variation of the action yields, after integration by parts:

$$\delta S(A) = -\int_{R\times\mathbb{R}} \delta A^{\nu}_{I}(\mathrm{D}^{\mu}F^{I}_{\mu\nu}) + \oint_{S\times\mathbb{R}} s^{\mu}F^{I}_{\mu\nu}\delta A^{\nu}_{I}, \tag{2}$$

where s^{μ} is the normal to the hypersurface $S\times\mathbb{R}$ in $M \times \mathbb{R}$. Now, for the first term of (2) to vanish for arbitrary variations of the gauge potential it suffices that the gauge potential satisfies the Yang–Mills equations. But the second term vanishes only if either (the normal component of) the field tensor vanishes along the boundary or δA^{I}_{μ} vanishes at the boundary. The first condition is severely limiting; the second is not a gauge-invariant condition.[4]

The standard approach to non-asymptotic spatial boundaries treats the lack of invariance of the subsystem by paring down gauge symmetries at the boundary. However, a recent flurry of papers on subsystems in gauge theory has called the pared-down treatment of internal boundaries of subsystems into question (see, e.g., Ramirez & Teh 2025, Geiller 2017; Donnelly & Freidel 2016, Gomes & Riello 2021, Riello 2021, Gomes 2019, Geiller & Jai-akson 2020, Carrozza & Höhn 2022 and references therein). New mechanisms, for instance, 'edge-modes', have been devised to maintain the gauge invariance of the internal boundary under all gauge transformations (e.g., even ones with support on the boundary). By edge modes, generally speaking, we mean sets of degrees of freedom that are localised on boundaries only, and particularly any physical reference frames that naturally can be constructed from them.

In the case of gravity, these degrees of freedom have received significant attention, starting implicitly with the study of asymptotic symmetries (He et al. 2014, Henneaux and Troessaert 2020), but further continued to surfaces at finite distances (Ciambelli and Leigh 2023, Freidel et al. 2020). A common theme is the presence of a set of symmetries on surfaces S of codimension 2, of the general shape

$$G_S = \mathrm{Diff}(S) \ltimes K^S \qquad K_{EH} = SL(2;\mathbb{R})_{\perp} \tag{3}$$

which contains diffeomorphisms of the surface acting on locally parametrised maps from S into a group K. In the case of the Einstein–Hilbert formulation of gravity, this group is given by shears and boosts in the 1+1D Lorentzian, normal plane of the surface S. There are many technical subtleties involved in the relation of these symmetries between finite and infinite surfaces, and the precise realisation of these symmetries on a system, for example in analogy to hydrodynamical systems, is the subject of ongoing study (Donnelly et al. 2020). Here, we want to highlight that a faithful action of these groups as symmetries requires the presence of a frame which is *not* a redundant part of the state of the system. Instead,

the boundary must be equipped with a frame of physical status, in order for the symmetry group to act properly on the system.

6. Edge modes provide/allow for the construction of physical frames

Using representational schemes consisting of physical reference frames or gauge-fixings, one can find gauge-invariant regional dynamical structures labeled by the (equivalence classes of the) fluxes at the boundary; see Riello (2021: Sec. 4) and Gomes & Riello (2021: Sec. 3). In a similar fashion, in Carrozza & Höhn 2022 and Carrozza et al. 2024 these dynamical structures are obtained by stipulating a representational scheme solely at the boundary; they call such conventions *boundary reference frames*.

Although these approaches vary to a certain extent, here we can try to describe their common core. The main idea is to find a physical representation of the symmetry-variant variables, which we called $h(\varphi) := \varphi^{\gamma(\varphi)}$, and correspond to the projection of the field along the corresponding gauge-fixing surface. Variation will then act not only on the bare variable φ, but also on the dressing, $\gamma(\varphi)$. This variation cancels the troublesome non-gauge-invariant boundary terms.

In a bit more detail, here is the rough idea of why this works (see especially Gomes & Riello 2017; Gomes 2019): The problem with the extra term in the variation of the action in (2) is essentially a lack of gauge-covariance under the variation δ. To correct for this lack of covariance, we can use analogies between finite- and infinite-dimensional geometry. Using these analogies, we can relate δ and spacetime derivative operators, d. And similarly, for theories with symmetries, there is an analogy between the infinite-dimensional space of models Φ and finite-dimensional principal bundles, P. In the finite-dimensional case, to get around a lack of gauge-covariance under spacetime-dependent gauge transformations, we introduce the connection ω on P through minimal coupling, i.e., so that, schematically, quantities acted on by the *horizontal* derivatives $\mathrm{d} \to \mathrm{d}_h := \mathrm{d} - \omega$ (where h stands for 'horizontal') remain covariant under spacetime-dependent gauge transformations. So too, in the infinite-dimensional case, we can justify the introduction of a connection-form ϖ, such that $\delta \to \delta_h := \delta - \varpi$ becomes a fully covariant operator. These 'connection-forms' are the infinitesimal versions of the dressing function $\gamma(\varphi)$. That is, the ϖ generalise the dressing of the fields that one would obtain via gauge-fixing to an infinitesimal setting; because they are (group-valued) functionals of the fields, they are relational constructions. Each choice of ϖ gives a decomposition of field variations in terms of 'pure gauge' and 'physical': but again, what is 'physical' depends on a choice.[5]

In a similar spirit, Carrozza & Höhn (2022) and Carrozza et al. (2024) demand a decoupling of the variational action principles of complementary spacetime regions separated by a timelike boundary (as in (2)). This is done in two steps: (i) restricting the state space by physical boundary conditions, such as fixing the electromagnetic field at a boundary, i.e., fixing a superselection, or, in their

words, a post-selected sector; and (ii) adding boundary terms to the action functional so that, within the physically restricted state space, the regional variational principles—such as in (2)—fully decouple. As we have described in this section, point (i) has a subtlety: in gauge theory, it is a nontrivial matter to fix physical, or gauge-invariant, boundary conditions. To accomplish that, Carrozza & Höhn (2022) and Carrozza et al. (2024) help themselves to reference frames at the boundary. Relative to these reference frames we can fix gauge-invariant boundary conditions for the gauge potential or the metric.

The resolution of the failure of gauge-covariance at the boundary is pursued differently in Donnelly & Freidel (2016) and follow-up papers (see, e.g., Geiller & Jai-akson 2020 for a more complete list). These papers add new degrees of freedom at the boundary with appropriate gauge-variance properties so as to cancel out the unwanted terms. In some circumstances the two approaches are related through a suitable interpretation of the new degrees of freedom (see, e.g., Riello 2021: Section 5; Regge & Teitelboim 1974: Section 5; Carrozza & Höhn 2022: Section 4; and Ramirez & Teh 2025). But many find the introduction of new degrees of freedom problematic (see Gomes & Riello 2021: Sec. 7; Riello 2021: 5.6; and Gomes 2019: Sec. 3.2 for critiques).

7. We are left with only physical reference frames, but a more realistic (and closer to operational) local physics

The end result of removing both kinds of idealisations, i.e., that of frames being expressed by coordinate choices and that of infinite spacetime regions, leaves us with models of gravitational systems defined on (collections of) finite regions of spacetime with geometric (thus gravitational) observables expressed as relations between field values (including the metric field).

Of course, it is difficult to explicitly realise such constructions. Classical deparametrisation of GR in terms of material frames has been achieved globally only in terms of not-entirely-physical material systems. A patchwise construction of generic gravitational dynamics with realistic matter frames is yet to be developed. Quantum reference frames, as studied so far, neglect entirely the gravitational coupling and most dynamical aspects. Constructions within full quantum gravity, thus both fully dynamical and fully quantum, have only been achieved in spatial approximations (Marchetti & Oriti 2022). Quantum-gravitational toy models, such as those employed in quantum cosmology in which relational dynamics is extensively studied, are indeed just toy models.

Despite all of the above limitations, thus, the lack of explicit, complete, and fully general realisations of a description of gravitational physics in a diffeo-invariant language entirely in terms of physical reference frames, the above arguments, and known partial results convince us that this is what we should strive for, to achieve a more realistic modelling of gravitational systems and spacetime physics. We now expand and draw some further implications from this conclusion.

8. Gravitational (spacetime) physics is intrinsically perspectival

As we have argued, a proper modelling of a (quantum-)gravitational system must invoke descriptions involving relations between subsystems, and such relations are expressed relative to physical reference frames used by some general kinds of observers. On the one hand, such a style of modelling is incentivised by the ambition to be realistic in our understanding of physics, but as we have argued, it is also made an implicit necessity by the formalities of gauge theories, in particular gravity. In general, now, these reference frames may be of totally different structure, validity, and quality. Therefore, their corresponding descriptions of a system are a priori very different as well.

To illustrate this idea, in its simplest form, we can point to the various different types of ways one can implement a physical clock. The, perhaps, most precise type uses atomic orbital transitions, but it is also possible to use something more elementary like a pendulum. In a gravitational context we can expect the two to behave qualitatively quite differently, as one of them makes explicit use of its gravitational interaction. However, in principle the atomic clock also couples gravitationally, and is bound to have a limited range of validity as a reliable reference frame. Therefore, the coupling of the reference frame and the remainder of the system is sometimes negligible, sometimes non-negligible, and sometimes even essential for its interpretation as a frame.

Now, according to a realistic account of agents and their physical reference frames, any set of relational observables is expressed, by definition, in the perspective of some agent participating in the dynamics of spacetime. Given this issue, one has already on the purely formal level an attachment of modelling activity and its results (predictions etc.) to the concrete perspective of some agent. Furthermore, existing models for these situations, such as different choices of physical clocks in quantum cosmology (Gielen & Menéndez-Pidal 2022, Giesel et al. 2023), show that even at a practical level, this attachment can be a nontrivial issue.

In a quantum context, there are even further complications beyond the classical regime. While classically, the relational strategy provides a resolution to the lack of diffeomorphism-invariant observables, their quantum versions are less useful in identifying conventional objects (like spacetime events) familiar from local non-relativistic physics (Kabel et al. 2024). Such complications then further emphasise that the choice of reference frame directly correlates with an appropriate set of accessible observables.

This correlation is particularly visible in recent works emphasising the role of physical clocks for algebras of gravitational observables in de Sitter spacetime (Chandrasekaran et al. 2023, Jensen et al. 2023, Vuyst et al. 2024). In de Sitter spacetime, it is common to have these algebras be trivial by themselves; however, upon inclusion of an observer in the form of a worldline equipped with a quantum clock of positive energy, one finds a nontrivial algebra of observables, which is even, in a precise sense, regulated by the quantum clock.

We, therefore, have to assert, when no further qualification is present, that concrete physical models associated with different physical frames are, in principle, unrelated. Thus, since the description is contingent on a physical frame whose qualities influence the modelling process, we can call it *perspectival*.

9. We may have no general covariance, not to mention invariance, in the more realistic, physical context

At this point, it should be clear that for any model of gravitational physics (as well as gauge physics), we have at our disposal, in principle, several possible physical frames. But the standard notion of general covariance (the standard independence of non-physical coordinate charts or diffeomorphism invariance) of general relativity is made irrelevant (because already accounted for) by the use of physical frames in the relational framework. Depending on their physical (dynamical) properties, including quantum features, one frame may be obviously more practically useful than another, but in a realist reading of the theory, this practical aspect would not point to a more fundamental ontological status for one frame over another.

The question arises, then, of whether we have any new notion of *physical* general covariance among the physical reference frames. That is, are there transformation maps between the different relational descriptions defined in terms of different physical frames, and what, if anything, is conserved or invariant under such mappings? One could argue that some form of physical covariance is expected, and we have both general mathematical arguments as well as explicit examples of such physical covariance, which are robust at the classical level. These examples usually involve using the redundant parametrised formulation written in terms of diffeo-covariant quantities only as a bridge between two different deparametrised formulations in terms of diffeo-invariant relational descriptions in terms of two different physical (material) frames. Such examples also usually involve going partially on-shell with respect to some equation of motion, limiting their generality and their applicability beyond the strictly classical context.

That said, even if full physical covariance could be achieved, one should not expect, generically, invariance of all physical properties or dynamical features under the identified maps between different relational descriptions corresponding to different physical frames. Such maps would not correspond to gauge symmetries and so we would lack any mathematical reason to enforce invariance. Again, we have examples pointing to these types of exceptions. As a relevant further complication, we have the issue that our understanding of general reference frame transformations at the quantum level is still a work in progress. Available results concern both simplified settings (Krumm et al. 2021; Vanrietvelde et al. 2023) as well as more generic constructions (de la Hamette et al. 2021). The possibility of switching between quantum reference frames in a full dynamical context passing through a (redundant) perspective-neutral formulation, like in the classical case,

is a promising avenue (Vanrietvelde et al. 2020; de la Hamette et al. 2021), and has been argued to be equivalent to, or even subsume other, more conventional approaches like the Page–Wootters formalism (Hoehn et al. 2021b, 2021a).

In the cases where such an understanding is under control, such as that of finite-dimensional quantum reference frames, we witness several salient, interesting features. Important properties of quantum systems like superpositions and entanglement of subsystems become, to a notable degree, perspectival. The same state of system and agents, as viewed from one particular agent's perspective, may be sharp or entangled in a certain observable, or it may be perfectly smeared out or disentangled in another agent's perspective (cf. Kabel et al. 2024 for examples). While this may simply hint at more complicated rules of quantum covariance (Cepollaro & Giacomini 2024), it is striking as it implies the sharpness of the frame—as linked, by the argument we made earlier, to the strength of backreaction—is further contextualised. Sharpness may be relative to the view of the system from an agent's perspective. Flipped on its head, taking the perspective of *any* one agent makes their reference frame (including any boundaries) sharp; but this then leads to a backreaction with possibly infinite strength due to fluctuations in conjugate variables. Therefore, quantum effects also put limitations on how strictly one can switch from one observer/agent's perspective to another.

At the end of these steps, we are thus left with more questions than answers, about what we obtain after removing all the mentioned idealisations from our models of gravitational systems and spacetime physics. The main conclusion is that we should be left with a description fully formulated in terms of relational quantities associated to our choice(s) of physical reference frames, and that we should strive to achieve such a description in greater generality, beyond the simplified or otherwise unsatisfactory examples we have currently at our disposal. We are also left with the fundamental questions of how to link the different descriptions associated to different choices of such physical reference frames, and of what, if anything, is invariant across them.

Discussion

In this contribution, we recalled ideas and results from three research directions, gravitational physics in finite regions, relational dynamics in classical and quantum gravity, and quantum reference frames, to advocate for a description of spacetime physics formulated fully in terms of physical reference frames: physical systems used as clocks and rods to localise other physical systems in space and time, and modelled without neglecting their physical features. We argue that this is the result of removing two types of idealisations from our physical models: that corresponding to idealised reference frames coded in coordinate systems or gauge fixings, and that corresponding to closed or asymptotic boundary conditions.

We emphasised that it is a partially open issue to understand what would be, in such description, a fully physical counterpart of the standard general covariance of gravitational physics with respect to changes of coordinate frames, i.e., a covariance with respect to transformations across physical reference frames. Moreover, we have emphasised that we have no reason to expect, even assuming that such physical covariance can be established, that physical properties would remain invariant under the same set of transformations. The dependence of the 'facts about any given gravitational system' on the chosen physical reference frame would instead be a feature to accept as fundamental and to understand in its full implications. The conclusion is that removing standard idealisations from gravitational physics makes a more fundamental (quantum) understanding of it necessarily perspectival, in the sense of being always relative to a particular choice of *physical* reference frame. This perspective is already present in a number of works in the literature and we emphasise its foundational importance.

Before we add some further comments about our arguments and conclusion, and their possible conceptual implications, we want to emphasise their generality. For concreteness, we have mostly focused on the gravitational case. Beyond its foundational significance and universality (any physical system is localised in spacetime, except spacetime itself, and interacts gravitationally), this case has the advantage of tying closely to our intuition. We have an intuitive understanding of what clocks and rods are, and what could model (and generalise) their functions. But most of our arguments and all the conclusions we have reached apply to any gauge theory. The recent work (and most results) about field theories in finite regions, edge modes, and boundary charges has been developed for both gravitational physics and any other gauge interactions, in particular non-abelian gauge theories. The role of edge modes in preserving gauge covariance in finite regions as well as their understanding as defining physical reference frames at the boundary of such regions remain valid. Moreover, any gauge symmetry can be construed geometrically (cf. Gomes 2024a), and the explicit inclusion of physical reference frames in our models is also necessary in gauge theories for gauge-invariant descriptions based on a relational strategy. The main difference is that such reference frames typically concern internal vector spaces, as opposed to spacetime, but this distinction does not affect the relevance of our arguments. These arguments appear, therefore, even more general than what we discussed.

Let us now discuss some more conceptual aspects of our conclusion about the importance of investigating further the issue of physical general covariance, and the dependence on any given choice of physical reference frame. These conceptual aspects, on the one hand, motivate further work linking this issue in the foundations of gravitational physics with other work in the foundations of quantum mechanics; on the other hand, they point to important open issues of a more general philosophical nature, that deserve as well further analysis and a proper treatment (which goes, however, beyond the scope of this contribution).

First of all, even if based on a large number of existing solid results, and regardless of how much we may be personally convinced, our conclusion is, to a large extent, mostly a pointer and an encouragement to further research. Each of the three research directions our argument builds on is a work in progress, with much left to be understood or established more rigorously. Some important open issues are: the quantum counterpart of the recent work on gravity in finite regions and edge modes; the definition and control over (relational) spacetime observables within quantum gravity approaches; the extraction of an effective gravitational dynamics within the same approaches; the extension of treatments of quantum reference frames to the dynamical gravitational context; the analysis of symmetries in field space that could represent (or contribute to) a physical counterpart of general covariance at both classical and quantum level, to name but a few.

Next, we argued that physical reference frames (and reference frames more generally) can be seen as a partial embodiment, in our physical models, of observers and of the very epistemic agents constructing/using such models. In the same direction, the removal of both kinds of idealisations, of idealised frames and of infinite or closed spacetime regions, which leads, we argued, to a description of gravitational systems that is necessarily relative to such physical frames, has to do with taking more seriously the role of observers and the open nature of physical systems. It is the role of observers and the system–observer split that have to become central in our understanding of physics; this is our conclusion, in more general terms, abstracting from our specific gravitational context.

From this more general perspective, there are other research directions that can be seen as supporting our general conclusion but which we did not mention. In particular, we have in mind the issues raised by the Wigner's friend scenario and its many modern generalisations, for example, the Frauchiger–Renner thought experiment (Frauchiger & Renner, 2018). In our reading, these scenarios suggest, in different ways, that indeed quantum physics may be intrinsically perspectival and that (quantum) facts are ultimately relative, in subtle and interesting ways (Brukner 2018, 2020). In turn, this aligns with a number of interpretations of the quantum formalism (pragmatist, relational, subjective, in different forms (Barzegar & Oriti 2024; Cabello 2017; Healey 2012; Healey 2023; Dieks 2022) that are based on relational, when not epistemic, approaches to quantum states and observables.

Following this chain of thoughts, thus taking seriously the link between physical reference frames in a gravitational context and observers (or epistemic agents), we can suggest further implications of our arguments, of an epistemological kind. The epistemological counterpart of (physical) general covariance can be taken to be complete intersubjectivity. By this, we mean the possibility of sharing fully, by appropriate translation across the perspectives associated to individual observers

or epistemic agents, the content of each perspective, not just approximately but exactly, not only part of the physical content of each perspective but all of it. In other words, different observers are able to recognise what, exactly, physical facts and properties are, with respect to each of them. The epistemological counterpart of complete invariance under change of (physical) reference can be taken to be a stronger notion of objectivity. By this, we mean a complete agreement, across observers (or epistemic agents), about physical facts and the properties of physical systems. In other words, different observers are not only able to recognise what, exactly, physical facts and properties are with respect to each of them, but they actually agree on such facts and properties.

In this vein, the classical and quantum-gravitational challenges to physical general covariance, which, we argue, are yet to be fully established and must be investigated in more detail because of their central importance, are then challenges to intersubjectivity. And the fundamental limitations we suggest for the existence of complete and exact invariance under changes of physical reference frames are then fundamental limitations from gravitational and quantum physics to objectivity.

Thus, with the above identifications in mind, our physical reasoning is inspired by and lends support to the many philosophical viewpoints (in the philosophy of physics and philosophy of science more generally) that have advocated for some perspectivalism in our epistemology, for a weakening of the notion of attainable objectivity, and in general for a greater role for agents and observers (some examples are Giere 2006; Massimi 2017; Brown 2009; van Fraassen 2008).

It is clear that, at the present stage, this conclusion is not much more than a(nother) pointer to further research, this time in the philosophical domain; and that proper philosophical analysis is called for, in order to put these suggestions on solid ground. We hope, however, that the foundational interest of such an analysis for physics and philosophy is also clear.

Acknowledgements

We thank one anonymous referee for several useful comments on the first version of this contribution. DO acknowledges support through the Grant PR28/23 ATR2023-145735 (funded by MCIN/AEI/10.13039/501100011033). SL thanks E. Telali for pointing out that single observables can still be made sharp in QRFs, B. Sahdo for the two examples of contrasting clocks and the Perimeter Institute for hospitality. Research at the Perimeter Institute is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MEDT. SL and DO acknowledge funding from the Munich Center for Quantum Science and Technology. HG acknowledges the British Academy for support.

Endnotes

1. The notions of agent and observer are obviously distinct, and modelling a system is not the same as observing it; however, in physics we model systems with the idea that their properties can be at least indirectly 'observed' or empirically accessed via their physical consequences, and in spacetime physics, which is our present focus, this 'access' to the system includes a determination of its location in space and time. Thus an epistemic agent modelling a system would also be 'observing' the system when specifying its location in space and time. Thus, in the following we often use either notion, depending on the aspects of the combined modelling/observing activity we want to emphasise.
2. Of course, in suitably engineered situations, one might devise a clock whose conjugate variable only indirectly couples to gravity. The general issue, however, remains.
3. This type of holism, or nonlocality, is a well-known issue for theories with elliptic initial value problems, e.g., Yang–Mills theory and General Relativity. For a reference that explores non-factorisability in the context of the holonomy formalism, see Buividovich & Polikarpov (2008). For the relation between different symmetries and locality in gauge theory, see Berghofer et al. (2023), and for the nonlocality of gravitational invariants see, e.g., Torre (1993), Donnelly & Giddings (2016); for more recent use of this non-factorisability in the black hole information paradox, see Jacobson & Nguyen (2019). For a discussion of the relation between the factorisability of Hilbert spaces and the augmentation of the phase space with 'edge-modes', which we will shortly discuss, see Ramirez & Teh (2025), Geiller & Jai-akson (2020), Carrozza & Höhn (2022).
4. It is important here that these are time-like boundaries; for the spacelike initial and final surfaces, one can implement whatever initial conditions one likes. For this reason, most other work deals with non-covariance in the Hamiltonian or symplectic formalism. And the boundary term gives rise to the symplectic potential: $\theta = \int E^i \delta A_i$, which defines the symplectic structure of the theory, $\Omega := \delta\theta$. This is done with much greater care in Carrozza & Höhn (2022, Secs. 4–7).
5. In the language of Rovelli (2002), they become *partial observables*. The subscript h is appropriate since ϖ is associated with a horizontal projection operator on Φ. The difference between a horizontal projection operator mentioned here and a projection onto a gauge-fixing surface is that the former only needs to be 'distributional': it acts as a projection only within each $T_\varphi\Phi$. When such a distribution is integrable, meaning that the associated infinite-dimensional connection form ϖ has no associated curvature, then the horizontal spaces foliate the space of models, and each leaf will correspond to the range of a particular choice of σ, and each initial value selects an entire leaf: this will be equivalent to a gauge-fixing. The infinite-dimensional connection-forms generically have curvature; but the curvature vanishes in the Abelian theory, i.e., electromagnetism (cf. Gomes et al. 2019 for a comprehensive account).

Bibliography

Bamonti, N. (2023), 'What is a Reference Frame in General Relativity?'. arXiv:2307.09338v3.

Bamonti, N., & Gomes, H. (2024), 'What Reference Frames Teach Us about Symmetry Principles and Observability'. arXiv:2410.12892v1.

Banerjee, K., Calcagni, G., & Martin-Benito, M. (2012), 'Introduction to Loop Quantum Cosmology', in *SIGMA* 8: 016.

Barzegar, A., Margoni, E., & Oriti, D. (2023), 'A Minimalist Account of Agency in Physics'. arXiv:2307.16054v1.

Barzegar, A., & Oriti, D. (2024), 'Epistemic-Pragmatist Interpretations of Quantum Mechanics: A Comparative Assessment', in *Foundations of Physics* 54: 66.

Berghofer, P., Francois, J., Friederich, S., Gomes, H., Hetzroni, G., Maas, A., & Sondenheimer, R. (2023), *Elements in the Foundations of Physics: Gauge Symmetries, Symmetry Breaking, and Gauge-Invariant Approaches* (Cambridge: Cambridge University Press).

Brading, K., & Brown, H. R. (2000), 'Noether's Theorems and Gauge Symmetries'. arXiv:hep-th/0009058v1.

Brown, J. D., & Kuchar, K. V. (1995), 'Dust as a Standard of Space and Time in Canonical Quantum Gravity', *Physical Review D* 51: 5600–29.

Brown, M. J. (2009), 'Models and Perspectives on Stage: Remarks on Giere's *Scientific Perspectivism*', in *Studies in History and Philosophy of Science Part A* 40: 213–20.

Brukner, C. (2018), 'A No-Go Theorem for Observer-Independent Facts', in *Entropy* 20: 350.

Brukner, C. (2020), 'Facts Are Relative', in *Nature Physics* 16: 1172–4.

Buividovich, P., & Polikarpov, M. (2008), 'Entanglement Entropy in Gauge Theories and the Holographic Principle for Electric Strings', in *Physics Letters B* 670: 141–5.

Cabello, A. (2017), 'Interpretations of Quantum Theory: A Map of Madness', in O. Lombardi, S. Fortin, F. Holik, and C. López (eds.), *What Is Quantum Information?* (Cambridge: Cambridge University Press), 138–144.

Carrozza, S., Eccles, S., & Hoehn, P. A. (2024), 'Edge Modes as Dynamical Frames: Charges from Post-Selection in Generally Covariant Theories', *SciPost Physics* 17: 048.

Carrozza, S., & Höhn, P. A. (2022), 'Edge Modes as Reference Frames and Boundary Actions from Post-Selection', in *Journal of High Energy Physics* 2022: 172.

Cepollaro, C., & Giacomini, F. (2024), 'Quantum Generalisation of Einstein's Equivalence Principle Can Be Verified with Entangled Clocks as Quantum Reference Frames'. arXiv:2112.03303.

Chandrasekaran, V., Penington, G., and Witten, E. (2023), 'Large N Algebras and Generalized Entropy', in *Journal of High Energy Physics* 4: 9, arXiv:2209.10454.

Chataignier, L., Hoehn, P. A., Lock, M. P. E., & Mele, F. M. (2024), 'Relational Dynamics with Periodic Clocks'. arXiv:2409.06479.

Ciambelli, L., and Leigh, R. G. (2023), 'Universal Corner Symmetry and the Orbit Method for Gravity', in *Nuclear Physics B* 986: 116053, arXiv:2207.06441

de la Hamette, A.-C., & Galley, T. D. (2020), 'Quantum Reference Frames for General Symmetry Groups', in *Quantum* 4: 367.

de la Hamette, A.-C., Galley, T. D., Hoehn, P. A., Loveridge, L., & Mueller, M. P. (2021), 'Perspective-Neutral Approach to Quantum Frame Covariance for General Symmetry Groups'. arXiv:2110.13824.

Dieks, D. (2022), 'Perspectival Quantum Realism', in *Foundations of Physics* 52: 95.

Dittrich, B. (2006), 'Partial and Complete Observables for Canonical General Relativity', in *Classical and Quantum Gravity* 23: 6155–84.

Donnelly, W., & Freidel, L. (2016), 'Local Subsystems in Gauge Theory and Gravity', in *Journal of High Energy Physics* 2016: 102.

Donnelly, W., Freidel, L., Moosavian, S. F., and Speranza A. J. (2021), 'Gravitational Edge Modes, Coadjoint Orbits, and Hydrodynamics', in *Journal of High Energy Physics* 9: 008, arXiv:2012.10367 [hep-th]

Donnelly, W., & Giddings, S. B. (2016), 'Diffeomorphism-Invariant Observables and Their Nonlocal Algebra', in *Physical Review D* 93: 024030.

Frauchiger, D., and Renner, R. (2018), 'Quantum Theory Cannot Consistently Describe the Use of Itself', in *Nature Communications* 9: 3711.

Freidel, L., Geiller, M., and Pranzetti, D. (2020), 'Edge Modes of Gravity—Part I. Corner Potentials and Charges', in *Journal of High Energy Physics* 11: 26.

Gaul, M., & Rovelli, C. (2000), 'Loop Quantum Gravity and the Meaning of Diffeomorphism Invariance', in *Lecture Notes in Physics* 541: 277–324.
Geiller, M. (2017), 'Edge Modes and Corner Ambiguities in 3d Chern–Simons Theory and Gravity', in *Nuclear Physics B* 924: 312–65.
Geiller, M., & Jai-akson, P. (2020), 'Extended Actions, Dynamics of Edge Modes, and Entanglement Entropy', in *Journal of High Energy Physics* 2020: 134.
Giacomini, F. (2021), 'Spacetime Quantum Reference Frames and Superpositions of Proper Times', in *Quantum* 5: 508.
Gielen, S., & Menéndez-Pidal, L. (2022), 'Unitarity, Clock Dependence and Quantum Recollapse in Quantum Cosmology', in *Classical and Quantum Gravity* 39: 075011.
Giere, R. N. (2006), *Scientific Perspectivism* (Chicago: University of Chicago Press).
Giesel, K., Li, B.-F., & Singh, P. (2021), 'Relating Dust Reference Models to Conventional Systems in Manifestly Gauge Invariant Perturbation Theory', in *Physical Review D* 104: 023501.
Giesel, K., Li, B.-F., & Singh, P. (2023), 'A Comparison of Different Choices of Clocks in a Reduced Phase Space Quantization in Loop Quantum Cosmology with an Inflationary Potential using Effective Techniques', in *16th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories* (World Scientific).
Giesel, K., & Thiemann, T. (2010), 'Algebraic Quantum Gravity (AQG). IV. Reduced Phase Space Quantisation of Loop Quantum Gravity', in *Classical and Quantum Gravity* 27: 175009.
Giesel, K., & Thiemann, T. (2015), 'Scalar Material Reference Systems and Loop Quantum Gravity', in *Classical and Quantum Gravity* 32: 135015.
Giulini, D. (2007), 'Some Remarks on the Notions of General Covariance and Background Independence', in *Lecture Notes in Physics* 721: 105–120.
Goeller, C., Hoehn, P. A., & Kirklin, J. (2022), 'Diffeomorphism-Invariant Observables and Dynamical Frames in Gravity: Reconciling Bulk Locality with General Covariance'. arXiv:2206.01193.
Gomes, H. (2019), 'Gauging the Boundary in Field-Space', in *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 67: 89–110.
Gomes, H. (2024a), 'Gauge Theory as the Geometry of Internal Vector Spaces'. arXiv:2404.10461v1.
Gomes, H. (2024b), 'Representational Schemes for Theories with Symmetries'. URL: https://philsci-archive.pitt.edu/24070/
Gomes, H., Hopfmüller, F., & Riello, A. (2019), 'A Unified Geometric Framework for Boundary Charges and Dressings: Non-Abelian Theory and Matter', in *Nuclear Physics B* 941: 249–315.
Gomes, H., & Riello, A. (2017), 'The Observer's Ghost: Notes on a Field Space Connection', in *Journal of High Energy Physics* 2017: 17.
Gomes, H., & Riello, A. (2021), 'The Quasilocal Degrees of Freedom of Yang–Mills Theory', in *SciPost Physics* 10: 130.
Gribov, V. N. (1978), 'Quantization of Nonabelian Gauge Theories', in *Nuclear Physics* B 139: 1–19.
Hausmann, L., Schmidhuber, A., & Castro-Ruiz, E. (2024), 'Measurement Events Relative to Temporal Quantum Reference Frames', arXiv:2308.10967.
He, T., Lysov, V., Mitra, P., and Strominger, A. (2015), 'BMS Supertranslations and Weinberg's Soft Graviton Theorem', in *Journal of High Energy Physics* 5: 151. arXiv:1401.7026

Healey, R. (2012), 'Quantum Theory: A Pragmatist Approach', in *The British Journal for the Philosophy of Science* 63: 729–71.

Healey, R. (2023), 'Quantum-Bayesian and Pragmatist Views of Quantum Theory', in E. N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University), Winter 2023 edition. https://plato.stanford.edu/archives/win2023/entries/quantum-bayesian/

Henneaux, M., and Troessaert, C. (2020), 'Asymptotic Structure of Gravity at Spatial Infinity in Four Spacetime Dimensions', in *Proc. Steklov Inst. Math.* 1: 309. arXiv:1904.04495

Hoehn, P. A. (2019), 'Switching Internal Times and a New Perspective on the "Wave Function of the Universe"', in *Universe* 5: 116.

Hoehn, P. A., Russo, A., & Smith, A. R. H. (2023), 'Matter Relative to Quantum Hypersurfaces', arxiv:2308.12912.

Hoehn, P. A., Smith, A. R. H., & Lock, M. P. E. (2021a), 'Equivalence of Approaches to Relational Quantum Dynamics in Relativistic Settings', in *Frontiers in Physics* 9: 587083.

Hoehn, P. A., Smith, A. R. H., & Lock, M. P. E. (2021b), 'The Trinity of Relational Quantum Dynamics', in *Physical Review D*, 104(6): 066001.

Hoehn, P. A., & Vanrietvelde, A. (2020), 'How to Switch between Relational Quantum Clocks', *New Journal of Physics* 22: 123048.

Jacobson, T., & Nguyen, P. (2019), 'Diffeomorphism Invariance and the Black Hole Information Paradox', in *Physical Review D* 100: 046002.

Jensen, K., Sorce, J., Speranza, A. (2023), 'Generalized Entropy for General Subregions in Quantum Gravity', in *Journal of High Energy Physics* 12: 020, arXiv:2306.01837

Kabel, V., de la Hamette, A.-C., Apadula, L., Cepollaro, C., Gomes, H., Butterfield, J., & Brukner, C. (2024), 'Identification is Pointless: Quantum Reference Frames, Localisation of Events, and the Quantum Hole Argument'. arXiv:2402.10267v1.

Krumm, M., Hoehn, P. A., & Mueller, M. P. (2021), 'Quantum Reference Frame Transformations as Symmetries and the Paradox of the Third Particle', in *Quantum* 5: 530.

Marchetti, L., & Oriti, D. (2022), 'Effective Dynamics of Scalar Cosmological Perturbations from Quantum Gravity', in *Journal of Cosmology and Astroparticle Physics* 07: 004.

Massimi, M. (2017), 'Perspectivism', in J. Saatsi (ed.), *The Routledge Handbook of Scientific Realism* (Routledge), 164–175.

Oriti, D. (2024), 'Hydrodynamics on (Mini)superspace or a Non-Linear Extension of Quantum Cosmology: An Effective Timeless Framework for Cosmology from Quantum Gravity', in S. De Bianchi, M. Forgione, & L. Marongiu (eds.), Time and Timelessness in Fundamental Physics and Cosmology; Historical, Philosophical, and Mathematical Perspectives (Fundamental Theories of Physics 216) (Cham: Springer), 221–52.

Ramirez, S., & Teh, N. (2025), 'Abandoning Galileo's Ship: The Quest for Non-Relational Empirical Signicance'. in *The British Journal for the Philosophy of Science* 76(2): 397–420.

Regge, T., & Teitelboim, C. (1974), 'Role of Surface Integrals in the Hamiltonian Formulation of General Relativity', in *Annals of Physics* 88: 286.

Riello, A. (2021), 'Symplectic Reduction of Yang–Mills Theory with Boundaries: From Superselection Sectors to Edge Modes, and Back', in *SciPost Physics* 10: 125.

Rovelli, C. (2002a), 'Partial Observables', in *Physical Review D* 65: 124013.

Singer, I. M. (1978), 'Some Remarks on the Gribov Ambiguity', in *Communications in Mathematical Physics* 60: 7–12.

Torre, C. G. (1993), 'Gravitational Observables and Local Symmetries', in *Physical Review D* 48: R2373–R2376.

van Fraassen, B. C. (2008), *Scientific Representation: Paradoxes of Perspective* (Oxford: Oxford University Press).

Vanrietvelde, A., Hoehn, P. A., & Giacomini, F. (2023), 'Switching Quantum Reference Frames in the N-Body Problem and the Absence of Global Relational Perspectives', in *Quantum* 7: 1088.

Vanrietvelde, A., Hoehn, P. A., Giacomini, F., & Castro-Ruiz, E. (2020), 'A Change of Perspective: Switching Quantum Reference Frames via a Perspective-Neutral Framework', in *Quantum* 4: 225.

Vuyst, J. D., Eccles, S., Hoehn, P. A., & Kirklin, J. (2024), 'Gravitational Entropy is Observer-Dependent'. arXiv:2405.00114.

Wallace, D. (2002), 'Time-Dependent Symmetries: The Link Between Gauge Symmetries and Indeterminism', in K. Brading and E. Castellani (eds.), *Symmetries in Physics: Philosophical Reflections* (Cambridge: Cambridge University Press), 163–173.

9
Relational Objectivity in the Presence of Finite Quantum Resources

Luis C. Barbado and Časlav Brukner

9.1 The Need for the Notion of Relational Objectivity

All our statements in physics are given in the form of propositions that use terms associated with classical concepts. This is not only a common practice but, according to Bohr, a necessity in order to understand and communicate to others what we have measured in the laboratory (Bohr 1958: 25–26):

> In this respect we must, on the one hand, realize that the aim of every physical experiment—to gain knowledge under reproducible and communicable conditions—leaves us no choice but to use everyday concepts, perhaps refined by the terminology of classical physics, not only in all accounts of the construction and manipulation of the measuring instruments but also in the description of the actual experimental results ...

To illustrate this let us consider the specific case of the spin projection measurement of a spin-one-half particle. A rather complicated experimental technique, in which a source of spin particles is prepared, the particles are passed through a Stern–Gerlach apparatus with the magnet aligned along, say the direction $\vec{n}$ (where $\vec{n}$ is the unit vector in the three-dimensional Euclidean space), and finally detected at one of the two detectors behind the magnet, can be summarised in a single sentence: *'The magnet was aligned along the direction $\vec{n}$ and the upper detector clicked.'* As such this sentence is a purely descriptive statement of what the experimental arrangement was and what was observed in a single run of the experiment. However, can we conclude from it that 'spin-up along the direction $\vec{n}$' is an objective property of the quantum system that exists prior to and independent of the experimental context?

According to textbook quantum mechanics, we assign a quantum state, formalised as a vector in a two-dimensional complex Hilbert space, to the spin particle. The state encodes the probabilities for the outcomes of all possible measurements that can be performed on the particle. This may suggest that statements

Luis C. Barbado and Časlav Brukner, *Relational Objectivity in the Presence of Finite Quantum Resources*.
In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press. © Oxford University Press (2026). DOI: 10.1093/9780198929277.003.0010

like 'The spin is up along the direction $\vec{n}$' refer to the particle *alone* as its objective property, independent of whether and which experiment is actually performed. However, from an operational perspective, the statement 'The spin is up along the direction $\vec{n}$' refers to the entire experimental context and is meaningful only if we have a way to account for $\vec{n}$ as a well-defined orientation in a three-dimensional space. In the laboratory, this is achieved by utilising *classical* physical devices that can serve as *practically infinite resources* of directional degrees of freedom, such as a macroscopic pointing arrow. Also, to complete the measurement and read out the result, we need a macroscopic device that determines whether the result was 'aligned' or 'anti-aligned' with the direction $\vec{n}$, such as a particle detector that 'clicks'.

Consider now a situation where the particle's spin is found to be up along the direction $\vec{n}$ after the measurement. If the measurement was non-destructive for the particle, the detection event can be used as a preparation for a subsequent measurement. If we want to measure the spin along another direction, say $\vec{m}$, it would be necessary to use another macroscopic pointing arrow. The quantum-mechanical calculus predicts that the probability p that the spin is aligned along the direction $\vec{m}$ is determined by the angle $\theta = \measuredangle(\vec{n}, \vec{m})$ between the first $\vec{n}$ and the second direction $\vec{m}$ of the macroscopic pointing arrows, as $p = \cos^2(\theta/2)$. We observe that all the statements and predictions that we were making, which initially appeared to refer to the quantum system alone, are ultimately about the relative configurations of the macroscopic devices as mediated by the (microscopic) spin-one-half system. Consequently, we conclude that the formal quantum-mechanical description of the system is *not*, in fact, describing any intrinsic properties of the spin in isolation. Rather, it is describing its properties in relation to the macroscopic devices that define the experimental context.

It can be argued that any physical arrow necessitates only a finite, albeit large, amount of resources to be specified. Therefore, the larger the set of values between 0 and π for the relative angles to be prepared between two arrows, the greater the amount of resources required to be specified. However, as for any macroscopic reference frame these resources exceed those of any finite-dimensional quantum system by far, we will consider the idealised limit of 'practically infinite resources', where classical relational configurations (such as a continuum of possible angles between physical arrows) emerge. The preceding discussions demonstrate that textbook quantum mechanics is inherently reliant on the use of these macroscopic reference frames in order to adequately describe the preparation and measurement of single quantum systems. In Section 9.3, we will introduce Penrose's spin network proposal as a potential formalisation of quantum theory that takes into account the finiteness of resources.

The necessity of introducing further physical systems, like the macroscopic arrows in the quantum theory of spins, in order to describe systems of interest is not a novel concept in physics. This is also evident in classical mechanics, where

the concept of *reference frames* plays an important role. These are also essential for providing physical meaning to, for instance, the orientation of the rotation axis of a gyroscope. Nevertheless, in the context of classical mechanics and classical statistical physics, this necessity, although present, can be ignored without compromising the physical description. This is because in these theories, one can assign a joint probability for the values of all possible measurement records for all possible measurements on the classical system. For example, any macroscopic reference frame for spatial directions allows the assignment of a joint probability to all components of the gyroscope's axis of rotation. As a result, the choice of macroscopic reference frame for preparing or measuring the gyroscope is arbitrary. Consequently, the identification of the components of the axis of rotation as intrinsic properties of the gyroscope within the formalism does not conflict with its ultimately relational nature.

The situation is fundamentally different in quantum mechanics. The no-go theorems of Bell (1964), as well as Kochen & Specker (1968), show that in general one cannot assign truth values, not even probabilistically, to the statements about measurement results for (actual as well as counterfactual) measurements prior to and independent of the experimental context—the entire classically describable measurement arrangement. There are basically two ways to respond to the no-go theorems. One can still decide in favour of retaining the classical notion of objectivity, in which one ascribes well-defined properties to a system, but which change when the experimental context changes. Bohmian mechanics is an interpretation of quantum theory that follows this view (Bohm 1952).

An alternative view is based on the understanding that at the core of the theorems of Bell, Kochen, and Specker is an *a priori assumption of separability between the system and the observational context.* It is only when this assumption is fulfilled that it is at all conceivable to define what the system is on its own. One can deny this separability and take the view that the context is a constitutive element that, jointly with the system, defines in relational manner what is objective in quantum theory.[1] This is the core idea behind the notion of *relational objectivity*: What is objective in quantum mechanics is not the intrinsic properties of the quantum system, but its specific relation with a *concrete physical macroscopic* context.

We notice that this change of perspective implies a relevant change in what a 'closed system' in quantum mechanics physically means: The notion of a closed quantum system in the 'textbook quantum formalism' would only be a shorthand notation for describing the physical relations between the given quantum system and the observational context. Important to notice, this change of perspective does not pretend that open quantum systems, in the canonical meaning of the term 'open', are more fundamental than closed quantum systems. Rather, it highlights that systems considered as closed in quantum mechanics are in general of

a nature that, from a strictly relational perspective, should be considered open. In particular, because the 'formal-theoretically closed' quantum system does not describe an *isolated* system, but its relations to macroscopic devices. And while, as discussed previously, in the classical case this lack of isolation has no further consequences, in the quantum case it is conceptually critical. For example, a spin-one-half particle is then never objectively oriented 'up' along the direction $\vec{n}$, but is objectively oriented along *this macroscopic red wood arrow here in the lab.* Given any other macroscopic, say, blue plastic arrow, the spin orientation is *undefined along the blue plastic arrow.* The probability that it will be aligned or anti-aligned along the blue arrow, after projection onto it, given that it was previously aligned along the red arrow, is given by an objective property *not of the spin particle, but of the macroscopic relation between the two arrows*, defined by the relative angle between their two directions. In this way, the absence of a joint probability distribution for all possible measurements on the spin is only natural, since the formalism never provides the probabilities for intrinsic properties of the spin, but rather for the results of a given *specific* experimental context. Any counterfactual reasoning about 'what unmeasured properties would have been if another measurement had been made' makes no sense if it is fundamentally impossible to separate the notions of system and context. Interpretations of quantum theory that are in agreement with this view include Quantum Bayesianism (QBism) (Caves et al. 2002), the relational approach (Rovelli 1996; 2005; Laudisa & Rovelli 2019), Healey's pragmatist approach (Healey 2012), as well as various interpretations that are either inspired by or are modern expressions of the Copenhagen interpretation (Bub 2017; 2020; Brukner 2017; Landsman 2017; Janas et al. 2022; Healey 2017; Cuffaro 2023; Auffèves & Grangier 2016).

In summary, we have exposed two critical features of quantum mechanics: the relational nature of any quantum state with respect to an experimental context, *and* the finiteness of the resources of any physical system, including the experimental context, with the textbook quantum-mechanical description being recovered in the limit of infinite resources of the context. Is there a description of quantum systems that incorporates, by construction, these two features? We approach Relational Quantum Mechanics as a natural candidate, and conclude that, while it embraces a relational description (as the very name indicates), it fails in taking into account the finiteness of the resources, which leads to several conceptual problems. We then explore the original construction of spin networks by Roger Penrose, which naturally incorporates *both* desired features. This leads us to conclude that such a construction is more adequate for the understanding of contextuality and non-locality in quantum mechanics, as well as the potential extension of the quantum framework beyond the conventional textbook quantum mechanics.

9.2 'Relational Quantum Mechanics' and the Problem of Insufficient Resources

Relational Quantum Mechanics (RQM) is an interpretation of quantum mechanics that takes a relational view of the properties of quantum systems as its fundamental premise (Rovelli 1996; 2005; Laudisa & Rovelli 2019). For RQM, properties of quantum systems are only meaningful in relation to other systems. In this sense, it is perfectly compatible with the theorems of Bell and Kochen and Specker.

In its attempt to extend the scope of the 'textbook quantum theory' to the domain outside the laboratory paradigm, where the notions of preparation, measurement, observation, context, and observer apparently play no role, RQM proposes relational objectivity between *any* two quantum systems, not only between a quantum system and a classically describable experimental setup. According to RQM, any interaction between two quantum systems should be considered as a kind of 'quantum measurement' in which the systems reveal to each other some *facts*, these facts being meaningful only in relation to the other system. In this manner, RQM also aims to resolve the ambiguity of the quantum measurement problem inherent to the laboratory paradigm. This ambiguity pertains to the question of which devices in the laboratory induce unitary evolution and which projections as a result of a measurement. In RQM, it is asserted that there is no ambiguity, as both types of evolution represent an interaction between quantum systems, and there is no fundamental distinction between the two.

Recently, however, several problematic aspects of the RQM programme have been raised in the form of no-go theorems (Pienaar 2021; Brukner 2021; Muciño et al. 2022). In particular, the theorem in Brukner (2021) demonstrates that RQM's notions of 'knowledge', 'information', or 'facts' that one quantum system may assign to or have in relation to another are incompatible with the fundamental features of the notions that are common in everyday life and physics praxis: (1) *the 'observer knows' the state of the 'observed system' when the states of the two are correlated*, (2) *distinguishable 'facts' are represented by orthogonal quantum states*.

We illustrate the essence of the no-go theorem with an example. Imagine that, from the point of view of C, two spin-one-half particles A and B are prepared into the singlet state—the totally anti-symmetric Bell state. According to RQM, the correlations between the two spins give rise to facts that A acquires relative to B and vice versa, of form 'spin B is "up" ("down") along direction z relative to spin A'. This is in agreement with condition (1). However, since *no basis is preferred for two spins*, one can rewrite the singlet state in an arbitrary but same basis of the two spins, and its form remains invariant. This leads to infinitely many further facts of form 'spin B is "up" ("down") along *arbitrary* direction $\vec{n}$ relative to spin A'. However, distinguishable facts like 'orientation "up" along direction

z' and 'orientation "up" along direction *x*', are *not* represented by orthogonal states, not even approximately, as long as they are represented by sufficiently low-dimensional systems such as spin-one-half particles. This contradicts (2).

The two fundamental features (1) and (2) of the notion 'knowledge' align closely with Bohr's quote from the beginning of the chapter. In the laboratory paradigm, the potential 'facts' that an observer can ascertain about a quantum system or that can be stored in a measurement device are distinguishable and thus represented by orthogonal (or nearly orthogonal) quantum states in the preferred basis induced by decoherence. This allows for the information about the facts to be read out unambiguously by multiple observers, communicated, and copied, ultimately leading to intersubjectivity about the facts among the observers. In contrast, the states between two correlated quantum systems are, in general, not orthogonal and cannot be read out unambiguously or copied—that is, 'qubits are not observers'.

The example provided for the no-go theorem already brings out the main problems into which RQM runs from a conceptual point of view: The central tenet of the RQM interpretation, namely the assumption that any interaction between two quantum systems can be regarded as acquiring certain facts of one quantum system about the other, in the form of outcomes of quantum measurements, is in direct conflict with the limited resources of microscopic quantum systems. These resources simply are insufficient to specify a quantum measurement. For example, in order for a quantum system to be able to measure the orientation of the spin of a spin-one-half particle, it should first and foremost be able to determine a direction and orientation in space with sufficient accuracy, which requires a large amount of directional resources. A magnet made up of a large number of equally prepared spin particles produces a magnetic field with a well-defined direction and orientation, which can then be used to determine a spatial direction along which other spin particles can be measured. But, for example, a half-spin particle alone is incapable of such a task because it alone cannot determine a direction in space. Assuming the opposite leads to contradictions as pointed out in (Brukner 2021). In addition, according to RQM a quantum system has the ability to act as a measuring device, manifesting 'facts' relative to this system, for which quantum-mechanical probabilistic predictions can be made. However, if we were to consider a simple quantum system that could act as a measuring device, we would face the same kind of questions that are common to the measurement problem: When, under what conditions, and in which basis is such a 'fact' realised? What is the state of the system after the fact is realised? What is the rate of 'probabilistic transitions' between past and new facts? We see that instead of solving the measurement problem, RQM installs it where it never existed, namely in situations where there are no measurements.

In their response to critics (Di Biagio & Rovelli 2022), the authors state that 'the information that a qubit has about another qubit need not be such that we can

understand what subjective experience of knowledge it would correspond to, and therefore RQM is under no obligation to solve the preferred basis problem'. But the no-go theorems never talked about 'subjective experience', they do not require the existence of an agent or even a measurement, they simply state that RQM's notions of 'knowledge', 'information', or 'facts' are incompatible with the very properties we use to identify these notions, if they were to be applicable to systems with limited resources. So it is RQM's task to explain what they mean, how we can operationally verify what they are supposed to mean, or to introduce new terms that avoid being misinterpreted as having anything in common with the conventional meaning of the terms.

Another counterargument given in Di Biagio and Rovelli (2022) is that 'RQM is about facts, not states'. According to this, the argument in Brukner (2021) would be mistaking the different relational descriptions of different systems, namely the information on facts about one qubit as revealed to the other qubit, and the information on the fact that the two qubits hold a relation as described by a third system. We will take up the argument in Brukner (2021) again, and reformulate it within the postulates of RQM itself (Rovelli 1996). We will show that the presumed distinction between facts and states is not tenable according to these postulates, and therefore show that the contradiction pointed out is still present and inherent in the RQM programme, and that it ultimately has to do with the lack of resources we argued in the previous paragraph.

Consider again the two spin-one-half particles A and B interacting with each other. As described by a third observer C, the particles are assumed to get entangled, again in the singlet Bell state; while for the systems themselves, A has revealed some fact about the other particle B, this fact being its spin along some direction. In terms of the postulates of RQM (Rovelli 1996), the act of interaction may be interpreted as A asking B the question 'Is your spin oriented along the direction $\vec{n}$?', where $\vec{n}$ is a unit vector in the three-dimensional space, and receiving a 'yes/no' answer. Note that, according to RQM, the set of questions that can be asked of a system does not depend on the resources the questioner has. This is the root of the problem: RQM tries to reproduce the structure of a quantum-mechanical measurement where there are not enough physical resources to do so.

According to Postulate 1 of RQM (Rovelli 1996) '[t]here is a maximum amount of relevant information that can be extracted from a system', and since in the case of spin one-half this maximum is one bit, according to RQM particle A already exhausted the maximum of relevant information about B. Next, according to Postulate 2, '[i]t is always possible to acquire new information about a system', but since A already has the maximal relevant information about B, the probabilities of the answers to any new question A wants to ask B are determined by the previous information in 'the best way' possible. At this point, we shall ask: What is the nature of this maximal information that A has about B, and that allows for determining the probabilities of future answers? Notice critically that this information

cannot be just one single 'yes/no'. One bit is meaningless without specifying the question to which the one-bit answer is given. The relevant information is that the question 'Is your spin oriented along the direction $\vec{n}$?' has been answered by 'yes/no'. So the total information consists of a pair, (yes/no, $\vec{n}$), the yes/no bit *and* the three-dimensional unit vector $\vec{n}$. Since having a spin aligned along the direction $\vec{n}$ is equivalent to having a spin anti-aligned along the direction $-\vec{n}$, formally (yes, $\vec{n}$) ≡ (no, $-\vec{n}$), the complete information about the spatial direction and the answer yes/no can be encoded in a vector of the Bloch sphere. Therefore, the space containing this information *must* be (isomorphic to) a projective two-dimensional complex Hilbert space. We conclude that the information on facts about B that A has must be represented by a quantum state.[2]

But, where is this information that A has about B physically stored? What the no-go theorems show is that there is no physical place in A to store the information on the fact/state about B that it has gained during the interaction in a reliable way. We notice that the necessity for information to be encoded in physical degrees of freedom, which is upheld by RQM, plays a pivotal role in the critique. In their response (Di Biagio & Rovelli 2022), the authors reject the criticisms on the grounds of a naturalistic view of information. However, we believe that their argument is actually put upside down. We quote (Di Biagio & Rovelli 2022: 12–13, emphasis in original, with some dummy labels changed to maintain consistency with the previous example):

> What is the meaning of the statement that *A* has knowledge about … *B*?
>
> There are two possible answers. The first is a naturalistic answer. The second is a dualistic or idealistic answer. According to the first, this is a statement about the actual physical configuration of the ink and the notebooks … in *A* and about the correlation of these with whatever can be observed in *B*. According to the second, *A*'s knowledge is something over and above its physical configuration. In this case, the 'inaccessibility' of *A*'s knowledge, namely of the 'universe as seen by *A*' is indeed there. But this only follows because one assumes that knowledge is unphysical.
>
> We adhere to a naturalistic philosophy. In a naturalistic philosophy, what *A* 'knows' regards physical variables in *A*. And *this* is accessible to *C*. If knowledge is physical, it is accessible by other systems via physical interactions.

We concur with the characterisation of knowledge according to the naturalistic and idealistic philosophies, and we acknowledge the choice of the naturalistic approach. However, we have demonstrated with a simple example that RQM, in its concrete attempt to reconstruct quantum mechanics, inadvertently falls into a picture closer to the idealistic philosophy. In particular, from what the example shows, it cannot be true that 'what A knows regards physical variables in A'. This is not possible as there are *no physical resources* in A where the physical information $\vec{n}$,

which is the fact/state that A gets to know about B, can be stored in a reliable manner that allows it to be used to determine the probabilities for subsequent facts about B in a quantum-mechanical manner. There are no ink or notebooks, as A is a simple spin-one-half particle. In even more straightforward terms, it is incorrect to assume that a spin-one-half particle can perform measurements on other spin particles as this task is beyond the capabilities of such a simple system.

Therefore, in our view RQM's attempts to evade statements about states as unphysical are merely rhetorical: the physical information in RQM looks and behaves like an ordinary quantum state. We notice that this conclusion was to be expected if RQM aimed to reproduce quantum mechanics in the limit of macroscopic preparation and measurement devices. There, quantum states are always conditioned on past facts about the preparation procedure. In this sense, for example, 'the spin-up along the direction $\vec{n}$' is a mathematical formalisation of the physical fact that the knob on the preparation device was oriented to produce particles with spin-up along the direction $\vec{n}$. Moreover, the quantum state represents also the information needed to compute the probabilities for the responses to the new question 'Is your spin oriented along the direction $\vec{m}$?' in the 'best way' possible, as Postulate 2 of RQM requires. Actually, this relation between facts and their representation in terms of quantum states was an integral part in (Brukner 2021), and so the counter-argument simply did not apply in the first place. It may only apply if by 'facts' something else is meant, which is what the no-go theorems prove. Alternatively, one could maintain the conventional meaning of 'facts'; however, this would result in contradictions of RQM with its own postulates.

To conclude this section, let us address the revised interpretation of RQM in Adlam & Rovelli (2022), where the absence of a preferred basis in correlations between quantum systems is not regarded as a defect but rather as a feature of RQM. It is proposed (Adlam & Rovelli 2022:15) that

> The alternative is to agree that quantum events do not typically have the simple form 'variable V taking value v relative to Alice'. Rather they must have a conjunctive form: 'variable V_1 taking value v_1 relative to Alice, and variable V_2 taking value v_2 relative to Alice, ...' and so on, specifying definite values for each of the variables ... in all of the different possible bases for it.

The question arises as to what the physical content and operational meaning of the statement that 'variable V_i takes the value v_i relative to Alice' for all possible bases i truly is. In the case of two spin-one-half particles A and B in the singlet state, these statements have an explicit form of 'spin B along direction x takes value "up" ("down") relative to A', 'spin B along direction y takes value "up" ("down") relative to A', etc. In order for these statements to have some explanatory or at least descriptive power, we require that they have a testable consequence when measurements are made, even though we acknowledge that the statements could be assigned to

the systems without the necessity of making the measurement. A natural way to establish a correspondence between the abstract statements and observations is to give to the statement 'spin B along direction x takes value "up" relative to A' an operational meaning that 'whenever B is measured to be "up" along x by a third system C, then A is found to be "down" along x by this same system C, if measured right after', in the spirit of the cross-perspective links proposed in Adlam & Rovelli (2022). However, this step brings us dangerously close to the hidden-variable trap of the Einstein–Podolsky–Rosen (EPR) programme (Einstein et al. 1935). For if one would require that the statements are 'real' for all directions $\vec{n}$ in the sense that their truth values are independent of whether and which measurement is performed on a distant system B, then there are 'elements of reality' in A in the EPR sense, and one is trapped in the hidden-variable hole. One can try to argue against this by negating the concrete operational correspondence or the reality of truth values. As one can deduce from the whole discussion, we would not know how to fix the first option without leaving the pivotal notions of 'fact' or 'quantum event' as empty terms with undefined observable consequences, and the second option is essentially the Copenhagen interpretation that RQM wanted to overcome.

9.3 Spin Networks as a Relational Description with Finite Resources

Spin networks, in their original formulation by Penrose (2010, 1971, 1972), constitute a framework in which the quantum theory of spins and the structure of the three-dimensional space of directions would be expected to arise together out of more primitive combinatorial rules. Since the aim is to reconstruct space rather than assume it, the combinatorial rules are based on the addition and subtraction of total angular momentum, which is rotationally invariant, does not single out any preferred spatial direction, and therefore does not require a pre-existing space.

Let us briefly review the basics of spin networks, in order to show how they could account for the notion of relational objectivity in standard quantum mechanics as emergent out of a more primitive notion of objectivity. A spin network consists of a graph in which line segments, called units, are identified with a non-negative integer each, which represents its total angular momentum in units of $\hbar/2$. The units come together in vertices, exactly three in each vertex. We may think of a vertex as representing the combining of two units together to make a third, or as the splitting of a single unit into two separate units. These combinations shall satisfy the rules for addition of quantum angular momentum, namely the triangular inequality and the conservation of fermions.

We provide an example of a spin network in Fig. 9.1. The units in the network that are not connected to two vertices are called 'free ends'. Given one network,

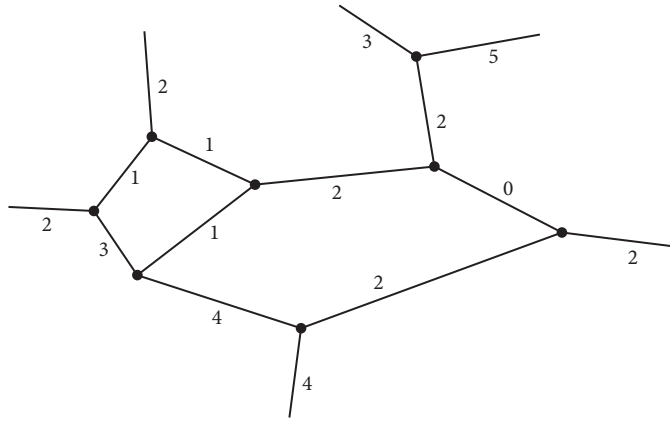


Figure 9.1 An example of a spin network. Line segments are called units, and are labelled with a non-negative integer that represents its total angular momentum in units of $\hbar/2$. Units meet in vertices, three in each, in a way that satisfies the rules of recoupling of quantum angular momentum.

one can consider the same network where two of the free ends merge together in a new vertex and form a new unit. The rules of spin networks provide a way to compute, through purely combinatorial means, the probabilities of obtaining the different possible values of angular momentum of this new unit (Penrose 2010).

The spin-geometry theorem (Penrose 2010, Moussouris 1983) shows that free ends with a large angular momentum can be used to define directions in three-dimensional space. These directions are however of an intrinsically *relational* nature, since what one can define is only the relative angle subtended by two free ends. So spin networks 'build up' themselves an emergent notion of space in the limit of large spins, rather than living in a pre-existing space.

The notion of relative angle between large free ends can be defined in the following way. Consider two free ends, with large values a and b. Consider a splitting of the first free end, where one of the split units is a spin-one-half particle (a unit 1), while the other unit has a value of $a - 1$. One can consider then the combination of the split spin-one-half particle with the second free end, and use the rules of spin networks to compute the probability p that the value of the unit resulting of this combination is $b + 1$. We depict the process in Fig. 9.2.

An interpretation of this procedure is the following: Since the spin-one-half particle, when split out of the original free end, decreased the angular momentum of it, we can say that such particle was 'prepared as aligned' with the first free end. If, when combined with the second free end, the particle happens to increase the angular momentum of it, we can say that the particle was 'measured to be aligned' with that other free end. Now, we can imagine for a moment that these free ends corresponded to strictly well-defined directions in space, through gyroscopes with high angular momentum. We know then from quantum mechanics that the

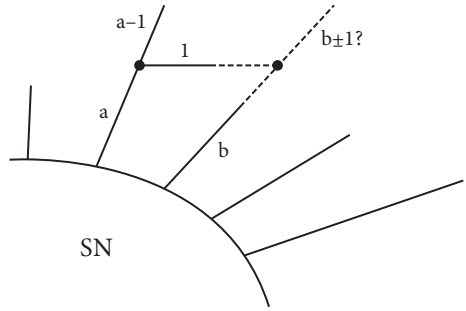


Figure 9.2 An example of a splitting of a spin-one-half particle from an end unit *a*, and its later coupling with an end unit *b*. The area denoted as SN stands for a generic spin network. The rules of spin networks allow to compute the probabilities that the value of the new unit formed is $b \pm 1$, given SN.

probability p that a particle aligned with the first gyroscope, when measured along the direction of the second, is found out as aligned with it, is given by the angle θ between the two angular momentum vectors of the gyroscopes, as $p = \cos^2(\theta/2)$.

In the case of spin networks, as we mentioned, the free ends do not correspond to any direction in a pre-existing space. So we do not compute p out of θ. We rather start with the probability p, as given by the rules of spin networks, as the fundamental quantity, and *define* θ from the relation above. We can in this way compute a collection of angles between any pair of free ends in a spin network. The spin-geometry theorem shows that, for spin networks with large end units, and such that the values of these angles remain stable enough under repetitions of the procedure described, this collection of angles satisfies the rules of spherical geometry, that is, of directions in a three-dimensional space, with arbitrary accuracy.

The spin-geometry theorem shows how spin networks give rise both to the rotational symmetry of the Euclidean three-dimensional space and the rotational degrees of freedom of quantum mechanics. From a formal point of view, this could be considered a trivial result, as the combinatorial rules of spin networks are inspired themselves by the standard recoupling theory of quantum angular momentum. But what we want to emphasise here is the conceptual twist that is suggested, considering the coupling mechanism and the probabilities as fundamental, and the space as emergent. Because it is this conceptual twist that can give rise, as an emergent notion in the limit of large units, to the *relational objectivity* that we discussed for quantum mechanics in the previous sections.

We propose to *identify* the elementary spin network, in the limit of very large spins and well-defined angles, with the elementary laboratory procedure where the preparation and measurement devices are oriented at a fixed relative angle; and to *identify* the split and coupling of smaller units between these devices (as the exchange described in the procedure above) as the quantum preparation and measurement processes.[3] Since, by the very nature of the framework, spin

networks provide the probabilities for only *one* experimental context, there is by construction no room for specifying joint probabilities for the outcomes of all possible experimental setups or for counterfactual reasoning, both of which are used to derive the Bell-like theorems. The large-enough spin networks constitute the *inseparable object–context entity* that is required in the relational approach.

We notice how the spin networks with finite units define probabilistic processes that may happen without giving a reference to a measurement. However, the operational significance of these probabilities is unclear. Under standard laboratory conditions, any experimental test to verify these probabilities would require a modification of the original finite spin network to include virtually infinite spin units to represent the infinite directional resources of the laboratory. Thus, any such test would only examine the compatibility of the Penrose framework with textbook quantum mechanics in the limit of infinitely large spin units. Nonetheless, together with the inherently contextual nature of the preparation and measurement processes in the construction, the other two key facts that we want to highlight about it are: (a) Unlike in RQM, the amount of possible relational configurations between systems scales with the resources of those. Therefore, the relational configuration between any two quantum systems, under the limitation of low resources, is not given by *any* possible quantum state of a given Hilbert space, but rather by a much more reduced range of possible relational configurations. This avoids the problems of RQM that we stressed in the previous section, which originated from the attempt to implement the structure of quantum measurements in the whole Hilbert space under the presence of finite resources. (b) The quantum-mechanical formalism for preparations and measurements, which assigns states in a Hilbert space and computes the probabilities out of them, would be recovered in the limit of large resources (the limit where it has been empirically tested), while at the same time the 'measurement devices', apart from their large size, have no special status compared to other spin units to which the remaining spin units are related. This prompts the question of the extent to which this offers a solution to the quantum measurement problem, which we leave for future research.

In order to fully reproduce quantum mechanics, accounting for preparations and measurements would not be enough. What about the other fundamental evolution in 'textbook quantum mechanics', namely unitary evolution? Although the original proposal of spin networks does not seem close to the possibility of reproducing a unitary evolution, we can find clear connections between spin networks and the recent results in Rudolph and Virmani (2005). There, it is shown that any unitary operation over a finite-dimensional Hilbert space H can be reproduced, to arbitrary precision, with the exclusive use of a sequence of *postselected* singlet/triplet measurements over pairs of qubits corresponding to partitions of a larger (but still finite-dimensional) Hilbert space H' that includes the original space $H \subset H'$, given an adequate initial state in H'. Since any sequence of

singlet/triplet measurements can be also reproduced as a trivial quantum angular momentum recoupling, we can conclude that one can also reproduce the effect of any unitary quantum gate to arbitrary precision with the use of a large-enough spin network (a spin network can be easily formulated as a linear transformation between finite-dimensional Hilbert spaces; see Major 1999).

It is not trivial, though, to visualise how a unitary evolution *in time* could be assimilated to an 'evolving' spin network, that is, a spin network in which different splits and combinations take place sequentially. We shall try nonetheless to give a *tentative* picture of what this could look like. Consider a given initial spin network. The units of a part of it, which we shall call the system *S*, will correspond to the inseparable object–context entity. These start to sequentially undergo processes of splits and combinations, enlarging the total spin network, that reproduce certain unitaries, as discussed previously. Now, let us consider that the part of the spin network that undergoes these internal processes is more complicated, and together with the system *S* we are interested in, contains something that can be called a clock *C*. The idea resembles the Page–Wootters mechanism (Page & Wootters 1983): (a) *C* undergoes processes in a specifically intertwined way, and in this sense in a synchronised way, with *S*; (b) *C* has a property, called time *t* (given for example by the relative angle between two large ends, acting as the hands of a clock), that grows uniformly, at least for a long-enough sequence of processes; and (c) this *t* is such that some external system, for example a physicist, can interact with *C* so as to read *t* without significantly disturbing the rest of the ongoing processes of *C* or *S*. Then, the physicist, if she happens to know the dynamics of *S*, could compute the probabilities of different results of a measurement that one could force within *S*, given a reading of the time *t*. These time-dependent probabilities could, in some macroscopic limit for *S* and *C*, reproduce a unitary evolution in time for *S* with arbitrary precision. We illustrate this evolution schematically in Fig. 9.3.

We shall make an important remark: Unlike the case of the measurement postulate, the unitary evolution in time of textbook quantum mechanics, due to its presentation in the form of a differential equation, makes an unavoidable use of the continuum *in each and every case* where it is applied. Therefore, *any* reproduction of unitary evolution in time with finite spin networks must be *necessarily* only approximate, even if with arbitrary accuracy. In the strict sense, this implies a departure from textbook quantum mechanics, although may not be distinguishable in the experimental range of application of quantum mechanics. Also, a departure from the formal unitary evolution of quantum mechanics does not necessarily mean a fundamentally irreversible evolution. Whether an evolution based on spin networks may destroy and create information at the fundamental level, or may rather have its own form of an information-preserving evolution, different from quantum-mechanical unitarity under finite resources, is an interesting question that goes beyond the discussion presented here.

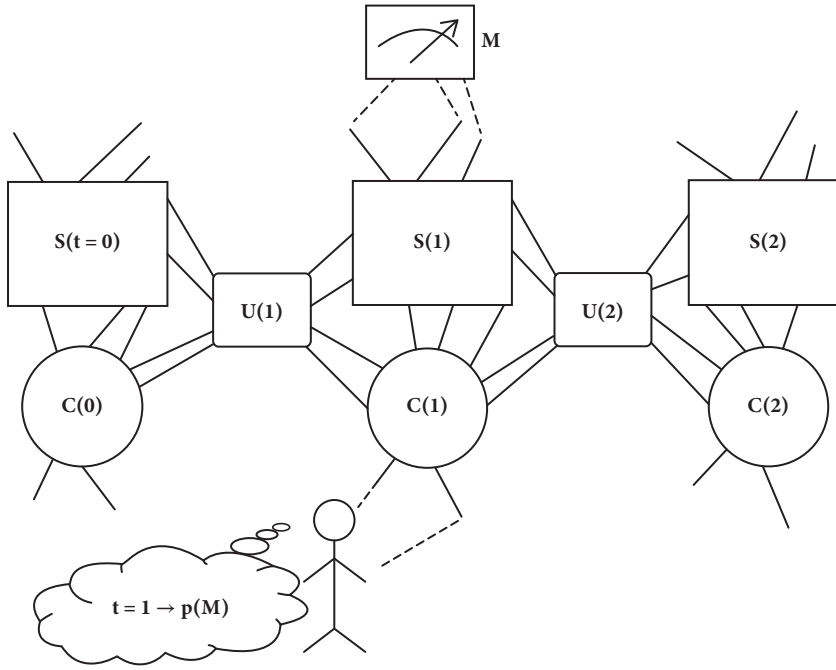


Figure 9.3 A schematic depiction of a *tentative* form of 'discretized approximate' unitary evolution in time, built up with spin networks. The areas denoted $S(t)$ and $C(t)$ represent spin networks that account for the system (object–context entity) and the clock, respectively, at different steps, and $U(t)$ represents a spin network evolving the system and the clock to the next step. An external agent could directly implement some measurement over the system with a measurement apparatus, obtaining some outcome M. This would, however, in general alter the system and interrupt the unitary evolution. But the agent can also access the clock and read only the time without significantly altering the system. This is information she can use, given that she knows the dynamics, to compute the probabilities $p(M)$ of different possible measurements and outcomes, which characterise the state of the system.

9.4 Conclusions and Outlook

All our statements and the language we use to communicate in everyday life contain implicit or explicit assumptions about infinite resources. When we say that we will meet someone at the Schottentor tram stop in Vienna at 2 p.m., we assume that the location in space-time is given with sufficiently large, practically infinite, resources. The statements we make about quantum systems are no different. When we say 'The spin of a spin-one-half particle is oriented "up" along the x direction', we are expressing specific laboratory arrangements and records in which there are virtually infinite resources for defining position and orientation in the three-dimensional space. These are the orientations of macroscopic devices,

such as the Stern–Gerlach magnets and detectors capable of producing detection events ('clicks').

We have argued here that 'textbook quantum mechanics' describes quantum systems in the presence of practically infinite resources, as is the case under standard laboratory conditions. Can one go beyond this? Relational quantum mechanics takes up the challenging task of making sense of quantum theory outside the laboratory, and while it adopts the relational perspective, it applies textbook quantum mechanics to systems with only finite resources, insufficient to make sense of statements like 'Spin B is oriented "up" along the direction x relative to spin A'. When A and B are two quantum systems of insufficient dimension there are no resources to define a spatial direction.

We propose here to keep the relational perspective but go a different way in taking the finiteness of resources as an integral part of quantum theory. We give an example of such an approach in terms of the original spin network construction of Penrose. The spin network is a framework in which one expects textbook quantum mechanics and space-time theory to emerge together from more primitive combinatorial rules and only in the limit of infinite resources (infinitely large spin units). Instead of statements like 'The spin is "up" along direction x', we can only make relative statements of the type 'the two spins are aligned' or 'the two spins are opposite', with the fundamental impossibility of specifying any spatial direction along which they are aligned, for the simple reason that in general there are not sufficient resources to define the directions. In contrast to RQM, these statements would not be made with respect to any presupposed basis or, equivalently, direction in space (since there are no resources to define one) nor to all of them simultaneously (since this would entail assuming hidden variables). Rather, these statements are made with respect to a rotationally invariant measurement of the total angular momentum of the two spins, with no mention of an external frame of reference.

It has been argued that the original construction of spin networks is formally capable of reproducing both the probabilistic assignments for measurement results and the unitarity of quantum mechanics, provided that sufficiently large networks are employed. From this perspective, textbook quantum mechanics can be seen as the limit of the spin network framework with infinite resources, which is the domain in which quantum pysics has been experimentally verified. The question of whether the spin-network construction has physical meaning under strictly finite resources, and whether it can be used to derive potential empirical consequences, is beyond the scope of this conceptual discussion. The aim of this discussion was to use spin networks to shed light on possible alternative approaches to fundamental problems in quantum mechanics, such as contextuality, objectivity, or relationalism.

In conclusion, the quantum-mechanical formalism inherently assumes a macroscopic context, with infinite resources, in the description of any purportedly closed quantum system. These systems would not be in true isolation but in

physical relation with the given context. We believe that, together with the spin networks formalism, works on *quantum reference frames* (Bartlett et al. 2007, Giacomini et al. 2019, Loveridge et al. 2018, Vanrietvelde et al. 2020), especially those that are not 'ideal' and hence do not presuppose infinite resources, may provide further clues as to how to achieve a relational formalisation of quantum theory in the presence of finite resources. What would a formalisation of the system–context entity look like, and how would the epistemic status of the entity be specified formally and in language? It is likely that this formalisation will move us further away from the visualisation so common in classical physics, which implicitly assumes infinite resources and is still somewhat present in textbook quantum mechanics through specific representations of a quantum state (as exemplified through statements such as 'The spin is "up" along the direction $\vec{n}$'), in much the same way that common language may lose its visualisation if the notions that require virtually unlimited resources, such as 'space', 'time', 'local', 'orientation', or 'distance' are omitted from the language. Whether this would require a revision of Bohr's dictum from the beginning of the paper remains to be seen.

Acknowledgements

The authors would like to thank Esteban Castro-Ruiz and Nathan Cohen for their invaluable input during the elaboration of this chapter. Additionally, they would like to thank Flavio Del Santo for his thorough reading and constructive feedback on the initial version of the text. This research was funded in whole or in part by the Austrian Science Fund (FWF) 10.55776/COE1 (Quantum Science Austria), [10.55776/F71] (BeyondC) and [10.55776/RG3] (Research Group 3). For Open Access purposes, the authors have applied a CC BY public copyright licence to any author-accepted manuscript version arising from this submission. This publication was made possible through the financial support of the ID 62312 grant from the John Templeton Foundation, as part of The Quantum Information Structure of Spacetime (QISS) Project (qiss.fr). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

Endnotes

1. While we will not discuss this further here, in our opinion, this does not require the presence of an agent or a laboratory; the experimental context could be defined by the prevailing type of decoherence in the environment of the system far from any agents or laboratories. However, even in such situations it might be useful to imagine fictitious agents that can actually use quantum theory to predict probabilities for individual results in a given experimental context or even choose different contexts.

2. As more complex finite-dimensional quantum systems can be built up from the composition of a finite number of qubits, and this way of encompassing larger systems with the formalism is explicitly embraced by RQM (Rovelli 1996), the argument provided applies to all the cases for which RQM has been explicitly discussed so far.
3. The identification of the recoupling of spins with larger spin systems as a measurement of the spin orientation is already present in previous works (Poulin 2006; Poulin & Yard 2007). However, in these works the identification is done within the full quantum-mechanical formalism (the resemblance to spin networks is only mentioned; Poulin 2006). While in such cases the resources required by the preparation and measurement devices may be explicitly present, the resources required for fixing their relative orientations are not, while in the spin-networks formalism they are through the existence of the rest of the spin network (SN in Fig. 9.2), which *prepares* the devices to hold a given angle.

Bibliography

Adlam, E., & Rovelli, C. (2022), 'Information is Physical: Cross-Perspective Links in Relational Quantum Mechanics'. arXiv:2203.13342.

Auffèves, A., & Grangier, P. (2016), 'Contexts, Systems and Modalities: A New Ontology for Quantum Mechanics', in *Foundations of Physics* 46: 121–37.

Bartlett, S. D., Rudolph, T., & Spekkens, R. W. (2007), 'Reference Frames, Superselection Rules, and Quantum Information', in *Reviews of Modern Physics* 79: 555–609.

Bell, J. S. (1964), 'On the Einstein Podolsky Rosen Paradox', in *Physics* 1: 195–200.

Bohm, D. (1952), 'A Suggested Interpretation of the Quantum Theory in Terms of Hidden Variables, I and II', in *Physical Review* 85: 166–93.

Bohr, N. (1958), *Atomic Physics and Human Knowledge* (New York: Wiley).

Brukner, Č. (2017), 'On the Quantum Measurement Problem', in R. Bertlmann & A. Zeilinger (eds.), *Quantum [Un]Speakables, vol. II: Half a Century of Bell's Theorem* (Springer), 95–117.

Brukner, Č. (2021), 'Qubits are Not Observers—A No-Go Theorem'. arXiv:2107.03513.

Bub, J. (2017), 'Why Bohr was (Mostly) Right'. arXiv:1711.01604.

Bub, J. (2020), 'In Defense of a "Single-World" Interpretation of Quantum Mechanics', in *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 72: 251–5.

Caves, C. M., Fuchs, C. A., & Schack, R. (2002), 'Quantum Probabilities as Bayesian Probabilities', in *Physical Review A* 65 022305.

Cuffaro, M. E. (2023), 'The Measurement Problem Is a Feature, Not a Bug – Schematising the Observer and the Concept of an Open System on an Informational, or (Neo-)Bohrian, Approach', *Entropy* 25: 1410. arXiv:2308.16371.

Di Biagio, A., & Rovelli, C. (2022), 'Relational Quantum Mechanics Is about Facts, Not States: A Reply to Pienaar and Brukner', in *Foundations of Physics* 52: 62.

Einstein, A., Podolsky, B., & Rosen, N. (1935), 'Can Quantum-Mechanical Description of Physical Reality be Considered Complete?', in *Physical Review* 47: 777–80.

Giacomini, F., Castro-Ruiz, E., & Brukner, Č. (2019), 'Quantum Mechanics and the Covariance of Physical Laws in Quantum Reference Frames', in *Nature Communications* 10: 494.

Healey, R. (2012), 'Quantum Theory: A Pragmatist Approach', in *British Journal for the Philosophy of Science* 63: 729–71.

Healey, R. (2017), *The Quantum Revolution in Philosophy* (Oxford University Press).

Janas, M., Cuffaro, M. E., & Janssen, M. (2022), *Understanding Quantum Raffles: Quantum Mechanics on an Informational Approach: Structure and Interpretation* (Springer).

Kochen, S., & Specker, E. (1968), 'The Problem of Hidden Variables in Quantum Mechanics', in *Indiana University Mathematics Journal* 17: 59–87.

Landsman, K. (2017), *Foundations of Quantum Theory: From Classical Concepts to Operator Algebras* (Springer International Publishing).

Laudisa, F., & Rovelli, C. (2019). 'Relational Quantum Mechanics', in E. N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab: Stanford University). Winter 2019 edition. https://plato.stanford.edu/archives/win2019/entries/qm-relational/

Loveridge, L., Miyadera, T., & Busch, P. (2018), 'Symmetry, Reference Frames, and Relational Quantities in Quantum Mechanics', in *Foundations of Physics* 48: 135–98.

Major, S. A. (1999), 'A Spin Network Primer', in *American Journal of Physics* 67: 972–80.

Moussouris, J. P. (1983), *Quantum Models of Space-time Based on Recoupling Theory* (University of Oxford).

Muciño, R., Okon, E., & Sudarsky, D. (2022), 'Assessing Relational Quantum Mechanics', in *Synthese* 200: 399.

Page, D. N., & Wootters, W. K. (1983), 'Evolution Without Evolution: Dynamics Described by Stationary Observables', in *Physical Review D* 27: 2885–92.

Penrose, R. (1971), 'Angular Momentum: An Approach to Combinatorial Space-Time', in T. Bastin (ed.), *Quantum Theory and Beyond: Essays and Discussions Arising from a Colloquium* (Cambridge University Press), 151–80.

Penrose, R. (1972), 'On the Nature of Quantum Geometry', in J. Klauder (ed.), *Magic Without Magic: John Archibald Wheeler, A Collection of Essays in Honor of his Sixtieth Birthday* (San Francisco: W.H. Freeman), 333–54.

Penrose, R. (2010), '31. Theory of Quantized Directions', in *Roger Penrose: Collected Works: Volume 1: 1953-1967* (Oxford University Press).

Pienaar, J. (2021), 'A Quintet of Quandaries: Five No-Go Theorems for Relational Quantum Mechanics', in *Foundations of Physics* 51: 97.

Poulin, D. (2006), 'Toy Model for a Relational Formulation of Quantum Theory', in *International Journal of Theoretical Physics* 45: 1189–1215.

Poulin, D., & Yard, J. (2007), 'Dynamics of a Quantum Reference Frame', in *New Journal of Physics* 9: 156.

Rovelli, C. (1996), 'Relational Quantum Mechanics', in *International Journal of Theoretical Physics* 35: 1637–78.

Rovelli, C. (2005), 'Relational Quantum Mechanics', in A. C. Elitzur, S. Dolev, and N. Kolenda (eds.), *Quo Vadis Quantum Mechanics? The Frontiers Collection* (Berlin, Heidelberg: Springer), 113–20.

Rudolph, T., & Virmani, S. S. (2005), 'A Relational Quantum Computer Using Only Two-Qubit Total Spin Measurement and an Initial Supply of Highly Mixed Single-Qubit States', in *New Journal of Physics* 7: 228.

Vanrietvelde, A., Hoehn, P. A., Giacomini, F., & Castro-Ruiz, E. (2020), 'A Change of Perspective: Switching Quantum Reference Frames via a Perspective-Neutral Framework', in *Quantum* 4: 225.

10
Open Systems and Autonomy

James Ladyman and Karim P. Y. Thébault

10.1 Introduction

The term 'open system' is widely used in science along with its contrary 'closed system'. It is often taken to be of great importance whether a system is 'closed' or 'open', and this feature may even be thought to be necessary, for a system to be studied by a particular science. Since Schrödinger puzzled about how living things can maintain their order and structure out of equilibrium, it has been observed that the Second Law of Thermodynamics refers to 'isolated' systems, while 'organisms are open systems' (Olby 1971: 128). The relevant sense of 'open' here involves the exchange of both energy and matter with the environment (von Bertalanffy 1950). Order and self-organisation can also emerge in nonliving systems that are open only in the sense of being driven by some force or flow of energy without exchanging matter with the environment. This is the sense in which complex systems are always open (Ladyman and Wiesner 2020).

Much of physics deals with closed systems (in various senses of the term distinguished and precisified below), but there are many models of open systems in physics, and open systems are of fundamental importance. For example, consider the significance of Brownian motion for the development of physics. There are many physical systems, including importantly stars, whose distinctive physical behaviour is only possible because they are open (Phillips 2013). Of course, all real physical systems are open in the sense that they interact or exchange matter or energy with their environments to some, even if only to a negligible, extent. However, the salient feature of the Brownian motion of a pollen grain, or the damped oscillation of a pendulum in the atmosphere of the Earth, is that interactions with the environment *non-negligibly* affect the system's dynamics. The 'scale-relativity' of the open–closed distinction as applied to concrete systems in scientific practice is an important part of the analysis of this chapter (and is the subject of Section 10.3). There are absolute distinctions between open and closed systems in the context of models and theories, however, these correspond to differences of degree in practice. Differences of degree add up to give the dynamics at different scales

James Ladyman and Karim P. Y. Thébault, *Open Systems and Autonomy*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press.
© Oxford University Press (2026). DOI: 10.1093/9780198929277.003.0011

studied by physics. The difference of degree between a few particles and 10^{23} is what constitutes gases and their emergent modal structure, for example, the gas laws.

The notions of open and closed in physics are entwined with ideas of autonomy, conservativeness, embeddability in a larger system, and formal properties of time evolution such as being completely positive, unitary, or measure-conserving. While the concepts of open and closed systems are of widespread foundational significance, the exact meaning of the terms 'open' and 'closed' varies between different contexts even within physics. To complicate matters even more, sometimes the term 'closed' is used to mean 'isolated', and this proliferation of concepts and language greatly complicates foundational and philosophical work. It is important to be clear about the relevant concepts and terminology so that fallacious reasoning based on conflation of the meanings of terms does not arise. Much of this paper designs and defends a conceptual framework for the disambiguation of scientific language regarding open and closed systems. Despite remaining very close to scientific practice, we arrive at some rather surprising conclusions. Most surprising of all, that the universe, which is usually taken to be a closed system, may be open in a sense relevant to classical cosmological models.

The next section analyses the open–closed distinction in more detail. We argue that although the distinction is made in different ways in different domains, its conceptual core is the conservation or not of some characteristic quantity. Section 10.3 discusses the importance of scale-relativity in how concrete systems are modelled as open or closed. Section 10.4 introduces the key idea of *autonomy* defined as a feature of a model of a system, namely that it contains well-posed equations of motion for the dynamical variables of the system alone. Autonomous models may nonetheless encode the dependence of a system's evolution on the environment, as with models of Brownian motion and many other models of open systems in physics. Hence, we further distinguish 'parameterised' autonomy, which involves parameters that represent features of the environment, from 'absolute' autonomy which does not involve such parameters. Section 10.4 also shows that autonomy of both types is independent of the open–closed distinctions of the next section. Autonomous open systems models are an important part of science in various domains. Section 10.5 considers the special case of open quantum systems and argues, following Cartwright (1983) and Cuffaro and Hartmann (2024), that there is no reason to take embeddability in a larger unitary system to be a fundamental feature of open-quantum-system models. Section 10.6 draws on the previous sections to argue that the argument that the universe as a whole must be subject to unitary time evolution conflates different senses of 'closed', and that it is in fact an open question whether we should model the universe as an open or closed quantum system.

10.2 Open and Closed

Two other foundational distinctions are required for our analysis. The first distinction is motivated by the observation that the same concrete system, for example, a particular pendulum in a particular laboratory, can be modelled as closed, if friction is ignored, or as open if it is not. Hence, it is important to distinguish between how the system is modelled and the system itself when discussing ideas of open and closed. In general, concepts and terms can be applied in the *material mode* to *concrete systems*, and in the *formal mode* to *models*.[1] Our usage of these terms is as follows. To discuss a physical system *X* in the material mode is to make an ontological claim about the properties of *X*. To discuss *X* in the formal mode is to refer to the representations (linguistic or mathematical structures) of *X*.[2] In what follows we clarify concepts and terms relating to open and closed systems by drawing attention to the formal mode/material mode distinction. Of course, any application of terms in material mode involves the conceptualisation of the system in some way. However, since the same material system may be modelled as an open or a closed system the material/formal distinction is required to say unambiguously what we mean when we say a system is open or closed. More generally we can obviously model real systems in many ways and consider many aspects of them while also idealising or abstracting, and sometimes the questions we ask are explicitly about the system as modelled.

The second distinction is that between a system and its environment. In material mode, unless the system is everything there is, the open–closed distinction always concerns the relationship between the system and its environment. However, in formal mode the environment may not be modelled at all and the system may be treated as if it is everything there is. For example, it may be stipulated that the system conserves or does not conserve a characteristic quantity without the environment or its effects on the system being considered at all. In scientific practice (except in cosmology), the environment is always the local surroundings of the system, not the whole universe. The immediate environment of the system is taken to be all that matters for the purposes of modelling the evolution of the system. The distinction between a system and its environment, surroundings, or other systems is usually predicated on there being some kind of boundary around the system so that whether the system is open or closed depends on what happens at the boundary. However, in some models there is no boundary, just the distinction between the system and its environment. Again, in formal mode there may be no environment since it may not be modelled at all, but of course in material mode every system except the whole universe has an environment. The local environment may be assumed to screen off any effects that the rest of the universe has on the system because they have to go through it, but of course if there is action at a distance this is not so.

The principal notions of open and closed systems applied in physical practice can then be distinguished as follows:

Definition 1 Closed Systems in Material Mode.

1. In the most general and strongest sense a system is closed (and isolated) iff it is not interacting with or affected by anything else.
2. In chemistry (and often in engineering) a system is closed iff it is contained within a boundary that no matter crosses. Heat or light can cross the boundary of chemically closed systems.[3]
3. In classical mechanics a system is closed iff it has no net (non-conservative) forces acting upon it (while also not exchanging matter with its environment).
4. In thermodynamics a system is closed iff it does not exchange heat or work with its environment (while also not exchanging matter with its environment).[4]
5. In statistical mechanics a system is closed iff the number of particles is constant and the probability flow is incompressible.
6. In quantum theory a system is closed iff its coherence does not change.

Obviously a system that is chemically closed need not be closed mechanically or thermodynamically, and indeed most chemical reactions produce or require heat. It is not true that a mechanically closed system must be isolated since a mechanical system may be closed whilst being subject to conservative forces. Indeed, the idea of a conservative force is central to our ability to assign a path-independent potential energy to a mechanical system moving within a force field, such as the gravitational field. A mechanically open system, which is acted on by non-conservative forces, is thermally open in that there is net work done on the system for any closed state space loop. In general, systems being isolated and systems being closed must be distinguished. An isolated system is a system that is not interacting with the environment or affected by anything else, so there is no exchange of matter, energy, information, or anything else across the boundary of the system. This may or may not be connected to failure of conservation since, for example, a system can be conservative whilst having boundary flows when these flows are net zero. In practice neither being closed nor isolated are absolute, but both can be effectively true of a system (as noted in Section 1 and discussed in the next section).

Clearly, the open–closed distinction is related to conservation such that in the context of particular physical theories there is typically a *characteristic quantity* such that the system is closed iff the quantity is conserved by its time-evolution, and otherwise open. A closed mechanical system is conservative with respect to energy, which is equivalent to there being no net non-conservative force since systems acted on by purely conservative forces are also energy conserving.

A closed chemical system is conservative with respect to mass. A closed statistical mechanical system is conservative with respect to the number of particles and probability current. A closed quantum system is conservative with respect to quantum coherence. Note that simple-harmonic motion that is conservative with respect to energy, and so mechanically closed, may be driven by an external conservative potential, and so not isolated.

All the above definitions can be understood in material mode (but only effectively unless they are applied to the whole universe), and when the conservation of the characteristic quantity is encoded formally within a model, they all have formal mode correlates as follows:

Definition 2 Closed Systems in Formal Mode.

1.* In the most general and strongest sense a model is closed (and isolated) iff it conserves everything that matters to the dynamics.
2.* In a chemically closed model mass is conserved, and so is the number of nuclei of each atomic species.
3.* In a model of a closed system in classical mechanics energy is conserved (the Hamiltonian is a constant of the motion).
4.* In a model of a closed system in thermodynamics heat and work are conserved.
5.* In statistical mechanics a model of a closed system has constant phase space dimension and volume measure.
6.* In quantum theory closed systems are those that undergo unitary time evolution (or whose state is in a decoherence-free subspace).

The formal mode and material mode descriptions may come apart. For example, a concrete engineering system may be modelled as conserving mass, and so as closed in the relevant formal mode sense, even though it leaks and so is not closed in the relevant material mode sense, and a concrete quantum system may be modelled as evolving unitarily, and so as closed in the relevant sense, even though it is subject to a small degree of decoherence and so is not closed in the relevant material mode sense. Systems that are not conservative are often assumed to be contained within a larger system that is conservative, for example, if the particles that leak out of a chemical system are present in the environment, they are conserved in the wider system. In the context of open quantum systems it is often assumed that they are embedded in larger closed quantum systems (as noted above and discussed in Section 10.6).

In foundational contexts the terms 'open' and 'closed' must be precisely specified because discussion often involves consideration of different theoretical domains in which different ideas of open and closed are at issue, and hence different quantities are conserved. For instance, it is fallacious to argue that a system

should be modelled as open in the formal mode sense of non-unitary time evolution, because it is interacting with another system, because that means only that it is not closed in the general material mode sense, unless some grounds are given to connect the different precisifications of the open–closed distinction. In fact, as discussed in detail in Section 10.6, it is not true that systems being affected by their environments cannot be modelled with unitary time evolution, so the general sense of open in material mode does not entail the formal mode sense of open in quantum physics.

In general, systems can be modelled by explicitly representing the environment in detail, or by only representing its effect on the system. Either way, the whole environment is never really modelled, but only the immediate surroundings of the system, so that together they can be thought of as a system that is effectively isolated from the rest of the universe. In this way, the dynamics of the system that depend on interactions with the environment can be modelled without modelling the whole universe, for example, by treating the environment as a heat bath, or as a source of random perturbations as in Brownian motion. The success of science in finding descriptions of effectively isolated systems, and open systems with local environments that are effectively isolated, makes it seem like the only way to understand open systems is as part of bigger closed systems. Similarly, it might be assumed that any open system can always be modelled as part of a bigger closed system, and since there is nothing more than the universe it might seem obvious that it must be a closed system in every sense. This assumption and the conclusion drawn from it are challenged in Section 10.6.

Many models work by assuming that open systems can be embedded in bigger closed systems, but in every such case it is always possible to make physical measurements to estimate the variables or parameters that represent the effect of the environment on the system, for example, the temperature of a heat bath, or viscosity of a liquid medium. In the case of the universe to embed it in anything bigger is speculative. The embeddability of open systems in larger closed and isolated systems is discussed further in Section 10.6. The next section considers how applying the open–closed distinction in any of its forms to concrete systems is scale-relative.

10.3 Scale-Relativity

We start with a simple observation regarding gravity and isolation that goes back at least to Russell (1903), cf. Barbour (2001), namely that gravitational interactions cannot be screened in Newtonian physics (nor in general relativity). Thus, every massive body in the universe interacts with every other massive body. Hence, in material mode strictly speaking *there is no such thing as an isolated sub-system* of the universe. Quantum field theory also implies that no system is ever really

isolated since the vacuum is a constant source of fluctuations that affect it. Furthermore, there is a fundamental ontological sense in which all systems studied in science are 'metaphysically open', namely that they are always interacting *to some degree* with other systems even non-gravitationally (unless the whole universe is modelled as per the discussion of Section 10.6). As Nielsen and Chuang (2010) put it '... in the real world there are no perfectly closed systems, except perhaps the universe as a whole' (353).

In the light of this observation it might seem that the material mode attribution of 'open' in terms of non-isolation loses its conceptual purchase since it applies to everything. The reason it does not is because of the general feature of scientific theorising, modelling, and practice called 'scale-relativity' by Ladyman and Ross (2007). For example, when we say that a flask of liquid is in thermal equilibrium with its environment this is implicitly relative to timescales that are short compared to the half-life of the liquid (which is evaporating and will eventually be entirely gone). Clearly, as Russell pointed out, all systems are materially open in the non-isolation sense because they are interacting gravitationally and otherwise to some degree. However, in many situations these interactions have negligible effects on aspects of their state. In such contexts they may be understood as isolated in the material mode since we can understand such statements in terms of *effective* isolation in accordance with scale-relativity.

Scale-relativity in the sense of approximation, or in the sense of restriction to some domain or regime, makes almost all of physical science concerned with effective rather than fundamental ontology (cf. Ladyman and Lorenzetti Forthcoming). Chemical bonds, fluids, and quasi-particles are obviously not fundamental entities but they are part of a domain, regime, or scale that is spatial, temporal, and energetic. It is difficult to overemphasise the significance of this observation for the interpretation of scientific language. In particular, when combined with the distinction between material and formal modes, the notion of scale-relativity renders transparent various aspects of scientific practice, that appear ambiguous or confusing otherwise. Most significantly, very often in science we find that *properties are represented as absolute in a model but when applied to a concrete system in material mode they are scale-relative*. For example, the smoothness of a surface is broken at small length scales, and the regularity of the oscillations of Newton's cradle is broken at long timescales. Being closed qua isolated (1. above) fits this pattern because systems are represented as isolated in absolute terms in a model, since exogenous interactions are completely excluded, whereas in material mode the system is effectively isolated in the sense that its dynamics are relatively unaffected by anything else over relevant timescales. Newtonian mechanics allows us to consider the solar system in just these terms: although it is subject to gravitational forces from the rest of the universe we can assume that its acceleration can be modelled as approximately uniform and linear and thus, via Corollary VI to the Laws

(Newton 1999: 423), treated as though it were not accelerating at all. This idealisation is justified precisely because effects arising from the non-uniformity of the gravitational field can be ignored at the scale in question. Furthermore, as noted in the previous section, even when systems are modelled as open, it is assumed that the system and its environment can be studied as if they form a larger system that is effectively isolated.

Scale-relativity is also, of course, significant for the understanding of the formal mode of open and closed discussed above—that is, the sense based upon conservation of a characteristic quality. When we talk of 'closed systems' in material mode we are referring to systems in which some relevant physical feature or quantity is conserved to some degree of approximation over the relevant timescale. For example, in practice the principle of conservation of mass is not exactly true of a chemical reaction, but in a large well-engineered chemical system any egress or ingress of matter is negligible compared to the size of the system and has no effect on the evolution of its overall properties such as, for example, the half-life of the reagents or the time when the reaction will effectively be over. Similarly, when we talk about concrete mechanical systems being closed in the material mode we are referring to the approximate conservation of the total energy of the system over the relevant timescale. A closed mechanical system has energy-conservative dynamical evolution, while an open mechanical system has dissipative dynamical evolution, where in each case the conservation and non-conservation is indexed to a timescale (or spatial scale). For example, the Earth–Moon–Sun system is understood as a closed mechanical system on the basis that the loss of energy is negligible on the relevant scales.

Conservation of the characteristic quantity may imply that interactions with the rest of the universe are not important to the system's dynamics at the relevant scale, as in a closed mechanical system, but it may not, as in the chemical case, where the exogenous application of heat to the system is relevant to its evolution. In general, interactions with their environment can be crucial to the dynamics of closed systems at the relevant scale, because there are situations where there is an exogenous driving force or a net-zero flux that is dynamically relevant even though the characteristic quantity is conserved. However, it is also often the case that a system can be understood as isolated on a given scale just because it can be understood as conservative of some quantity at that scale. It is appropriate to talk of 'open systems' in physics in material mode when referring to concrete systems in which the relevant physical feature or quantity is not conserved over the relevant timescale.

As noted above, isolation and non-isolation, and conservation and non-conversation are typically represented as absolute in a model when in material mode they are scale-relative. That is, a system is represented as isolated or as

conservative of the relevant quantity for the purpose of modelling, because at the scale picked out by the modelling context the evolution of the system is as if it was isolated, or the property is approximately conserved. Typically the scales in question are timescales, but spatial and numerical scales, and macro- and coarse-grained states, may also be relevant. For example, the individual particles in a gas are affected by the motions of atoms in other solar systems, but the evolution of the aggregate properties of the gas such as its pressure and temperature are not.

For mechanical systems the formal mode correlate of being closed qua conserving energy is typically found in terms of a Hamiltonian representation with both a privileged conserved function, the Hamiltonian, and a conserved phase space measure, the Liouville measure. For mechanical systems the formal mode correlate of being an open qua non-conservative system is typically found in terms of a contact Hamiltonian representation with non-conservation of the Hamiltonian, and a non-conserved phase space measure (Bravetti and Tapias 2015, Bravetti et al. 2017, Bravetti et al. 2020). A vivid analogy can be made between conservation of the phase space measure and the flow of an incompressible fluid and, in turn, between non-conservation of the phase space measure and compression of a phase space fluid, hence the term 'measure compression' is used to talk in formal mode about open systems (one could, of course, also have measure expansion under open-system dynamics).

In experimental contexts it is often the case that scientists select and prepare physical target systems so that they can be represented as isolated and conservative on the scale relevant for the experimental inferences they wish to draw. Thus, experimental scientists often select and prepare physical target systems with scale-relative ontological features that make them closed or isolated in the relevant sense.[5] The important point is that such a target–model relationship is only one part of scientific practice. There are also, however, examples where scientists model systems they select and prepare as open, as in experiments on decoherence or cooling. For real systems, the criterion for being open or closed is always whether the quantities in question, are conserved or not conserved. Conservation is scale-dependent because in real systems it is always approximate, and because the system is studied at a particular scale or scales. Interactions may be ignored in an effective description because they have a negligible effect or because they are irrelevant to some aspects of the system and their dynamics.

Scale-relativity applies to other features of systems too. For example, in quantum information theory it is possible to have a model of the effect of noise in a channel at one scale that is Markovian, even though at another scale the evolution is not Markovian (Preskill 2018). The next section considers general ways the dynamics of both open and closed systems can be modelled with a focus on autonomy as a feature of the representation of dynamics within the model.

10.4 Autonomy

The concept of *autonomy* goes back to its original ancient Greek coinage in the play *Antigone* by Sophocles.[6] The eponymous heroine is about to be led away to be buried alive for transgressing a law laid down by Kreon, King of Thebes, forbidding the burial of her brother. By way of a rather odd consolation, the chorus sing to her:

> Is it not with fame and praise that you depart to the corpses' depths? You were not struck down by wasting sicknesses, nor did you pay the wages of the sword, but autonomous [*αὐτόνομος*] you alone of mortals go living into Hades. (McNeill 2011: 412)

The implication is that Antigone is 'autonomous' in the sense that she, rather than the king or the gods, is the author of the laws of behaviour that she obeys in the act of burying her brother. Thus to be autonomous, in the Sophoclean sense, is to be the creator of your own laws: *auto-nomos*.

Similarly, in science and the philosophy of science the word 'autonomy' connotes that a system in some sense has its own laws, and that they are natural rather than socially constructed or supernatural. In particular, the notion of the autonomy of higher-level laws from lower-level laws is prominent in the context of the reduction/emergence debate. For example, we might understand the autonomy of higher-level laws as a criterion for genuine emergence and explicitly characterise autonomy in terms of the coarse-grained dynamics being independent of micro-dynamics (Franklin 2020, Robertson 2020, Palacios 2022). Paradigmatically, in continuum fluid mechanics there are autonomous laws for macro-scale variables independent of molecular-scale dynamics (Darrigol 2013, Batterman 2018).

The notion of autonomy developed here is related but more basic. We are interested in a model's capacity to produce independent and suitably behaved solutions to the equations of motion. Hence, we understand autonomy in formal mode and define it as follows:

Definition 3 Autonomy: a model of a physical system is *autonomous* iff it includes no explicit *dynamical variables* that encode degrees of freedom other than those of the system, and there are *well-posed dynamical equations* for these variables.

This definition of autonomy allows that a model of a system can be autonomous even when boundary or initial conditions must be specified, cf. Sloan (2023). This is sensible since boundary or initial conditions are always needed in practice. It also allows us to dismiss putative failures of autonomy that result purely from improperly specified boundary conditions.

In material mode a concrete system is autonomous to the extent that its dynamics can be modelled as *effectively* autonomous. We thus have that when

we talk about a *system* being autonomous we are making scale-relative statements about it that typically relate to how a system can be modelled in some respects to some degree of accuracy over some timescale.

There are three ways for a model not to be autonomous. First, and most obviously, the model may include dynamical variables that encode degrees of freedom other than those of the system. For example, a model of the Earth–Moon system that also includes the sun is not an autonomous model of that system both in the trivial sense that it includes a dynamical variable corresponding to the sun, and in the non-trivial sense that there is no well-posed sub-equation within the model for the Earth–Moon variables on their own. The second way for a model to be non-autonomous is if there are only variables for the system's degrees of freedom, but the equations of motion are not well-posed due to the presence of underspecified functions, leading to an underdetermined system of equations. For example, a model of a banknote being dropped off a building can be written in terms of variables for the banknote together with a conservative gravitational force and a non-conservative damping force. However, the damping force term will involve an underspecified function varying over space and time, and this means that the equations of motion are underdetermined.[7] The third way in which a model may fail to be autonomous is via the failure of the system of equations to be well-posed due to a formal breakdown of integrability such as a collision singularity in gravitational *n*-body dynamics (Saari 1973).

Lack of autonomy might sometimes be due to openness, but in general openness is compatible with autonomy because there are frameworks for constructing autonomous models of open classical and quantum systems (Breuer and Petruccione 2002, Sloan 2018). Autonomy and being closed or isolated dissociate, and there is good reason to keep them separate in discussion of open systems (there are examples to illustrate this point below). It is thus not always the case that autonomy can be restored by simply enlarging the boundaries of the system being represented. Autonomy is an analytic property of a system of equations relating to the generation of determinate evolution of the variables describing a system. Its failure may or may not be related to the model neglecting couplings to other systems. A model of a system with the dynamics including parameters that represent the effects on the system of an environment or bath is autonomous in the sense above. Hence, as mentioned above, autonomy is compatible with the system being open in various senses.

If autonomy is defined in this way there are two substantively different ways models can be autonomous:

Definition 4 Two Types of Autonomy.

Absolute Autonomy (A-Autonomy): There are well-posed dynamical equations in the model that are completely independent of any further systems.

Parameterised Autonomy (P-Autonomy): There are well-posed dynamical equations for the system but the dynamics depends on parameters or non-dynamical variables that encode features of other systems.

If the trinary distinction between non-autonomy, parameterised autonomy, and absolute autonomy is combined with the binary distinction between open (qua non-conservative of some characteristic quantity) and closed (qua conservative of some characteristic quantity), there are six possible categories:

	Non-Autonomous	**P-Autonomy**	**A-Autonomy**
Open	Open Non-Autonomous	Open P-Autonomy	Open A-Autonomy
Closed	Closed Non-Autonomous	Closed P-Autonomy	Closed A-Autonomy

Each category is explicitly defined and illustrated with an example from classical mechanics below.

Definition 5 Six Categories of Dynamical Model.

Open, Non-Autonomous: The characteristic quantity is not conserved and there are not well-posed dynamical equations for the dynamical variables of the model. Classical Mechanics: banknote system modelled mechanically by a point particle in a Newtonian gravitational potential field with an underspecified damping function.

Closed, Non-Autonomous: The characteristic quantity is conserved and there are not well-posed dynamical equations for the dynamical variables of the model. Classical Mechanics: n-body Newtonian gravitational system with collision singularities.

Open, Parameterised Autonomy: The characteristic quantity is not conserved and there are well-posed dynamical equations for the dynamical variables of the model such that the dynamics depends on parameters or non-dynamical variables that encode the effect of other systems. Classical Mechanics: damped oscillator model in Newtonian mechanics with external potential (pendulum).

Closed, Parameterised Autonomy: The characteristic quantity is conserved and there are well-posed dynamical equations for the dynamical variables such that the dynamics depends on parameters or non-dynamical variables that encode the effect of other systems. Classical Mechanics: undamped oscillator model in Newtonian mechanics with external potential (pendulum).

Open, Absolute Autonomy: The characteristic quantity is not conserved and there are well-posed dynamical equations for the dynamical variables of

the model that are completely independent of any further systems. Classical Mechanics: damped oscillator in Newtonian mechanics with internal potential (spring).

Closed, Absolute Autonomy: The characteristic quantity is conserved and there are well-posed dynamical equations for the dynamical variables of the model that are completely independent of any further systems. Classical Mechanics: undamped oscillator in Newtonian mechanics with internal potential (spring).

The above analysis of the open–closed distinction and the two types of autonomy shows that open systems can be and often are modelled as autonomous. Hence, being closed should not be conflated with being modelled as autonomous, however, it might still be expected that autonomous open-system models are ultimately reinterpretable in terms of parameterised autonomy. If so then, while there may be open-system models with well-posed dynamical equations of their own and no explicit parameterised dependence on other systems, these models should always be understood as *implicitly* representing further systems via a parameter. This would amount to insisting that our example of a damped oscillator in Newtonian mechanics with internal potential is only *apparently* absolutely autonomous because the spring potential has its own microphysical degrees of freedom, whose inclusion would 'close' the model leading to energy conservation, or at least making the model autonomous only in the parameterised sense. Section 10.6 returns to the status of absolute autonomy in open-system models in the context of cosmology.

The next section considers the interrelation between openness and the different forms of autonomy in the context of open quantum systems, and considers properties of dynamical maps that further complicate the investigation.

10.5 Open Quantum Systems

In the context of quantum physics it is important to distinguish quantum-open from the other notions of open described in Section 10.2. Many authors move back and forth between the formal mode notion of quantum-open, and material mode notions of closed systems having to do with being isolated or not interacting with the environment. For example, Nielsen and Chuang (2010) say that 'the dynamics of a closed system are described by a unitary transform' (357), but they frame this statement with reference to whether or not the system is interacting with the outside world. However, as noted above, there are many unitary quantum models of systems that are subject to effects from the outside world, as when particles are confined in a potential well, and there are many unitary models of systems that lose

energy (and so have a time-dependent Hamiltonian); cf. Breuer and Petruccione (2002: 110). Furthermore, it is possible to coherently drive quantum systems so that they move over macroscopic distances (Alberti et al. 2009). In general, any application of quantum mechanics in which the effects of electromagnetism are encoded in a potential cannot describe a system that is isolated in the material mode since the potential comes from the environment of the system, and this is orthogonal to the question of whether or not the model of the system has a unitary dynamics. It is also possible to model isolated systems with non-unitary dynamics in contexts where the Hamiltonian is not essentially self-adjoint. Furthermore, there can be decoherence without dissipation (Unruh 2012). Hence, quantum-open is not necessary or sufficient for openness in the other senses of Section 10.2. Note also that it is not the strength of interaction that determines whether it requires a quantum-open model, and that there are open-quantum-system models with a classical environment (Kapral 2015).[8]

Let us consider some important formal properties of unitarity in relation to probability, purity, entropy, and well-posedness. First, most straightforwardly, unitarity is sufficient but not necessary for probability conservation, since there are non-unitary dynamical equations that preserve probability. Unitary time evolution conserves the norm induced by the inner product. In particular, unitary time evolution is sufficient (although not necessary) for the preservation of the *purity* of a quantum state when represented as a density operator, ρ, which is a positive semi-definite operator of unit trace. Explicitly, the purity of a state is given by $\gamma = \mathrm{Tr}(\rho^2)$. Pure states are such that $\gamma = 1$. The quantum operation $\Lambda(\rho) = U\rho U^\dagger$ preserves the purity of ρ *if* U is a unitary operator (Jaeger 2007). Purity preservation is equivalent to the conservation of linear entropy since $S_L = 1 - \gamma$. There is also a connection between unitarity and the conservation of the informational or von Neumann entropy $S = -tr(\rho \ln \rho))$. In particular, the von Neumann entropy is also invariant with respect to the quantum operation $\Lambda(\rho) = U\rho U^\dagger$ since $S(U\rho U^\dagger) = S(\rho)$ (Breuer and Petruccione 2002: §2.3). Hence, unitary dynamics conserves both the linear and von Neumann entropies, and in so far as they are taken to be genuinely thermodynamical, the connection between unitary dynamics and being closed in the thermodynamic sense is a natural one.

There are strong results connecting unitarity with well-posedness. In particular, even for time-dependent Hamiltonians a general condition on the *boundedness* of the Hamiltonian implies the existence of a unitary propagator, which in turn implies that a quantum model will necessarily have well-posed equations of motion. Essentially, the Hamiltonian is restricted to a strongly continuous map from the real numbers into bounded self-adjoint operators on a Hilbert space. This then allows the definition of a Dyson expansion which in turn implies the existence of a unitary propagator. The continuity properties of the propagator are sufficient to prove the existence of well-posed equations of motion (Reed and Simon 1975: Theorem X.69). There is thus a strong formal connection between well-posed

dynamics and unitarity. However, the converse does not hold, and there are models with non-unitary dynamics via master equations that have well-posed dynamical equations.

To understand the contrast between open and closed quantum models consider first the paradigmatic equation for a unitary quantum dynamics of density operators, the von Neumann equation:

$$\dot{\rho} = -i[H, \rho], \tag{1}$$

where the evolution of density operator ρ is unitary provided the Hamiltonian H is self-adjoint. The paradigmatic equation for open quantum systems is the Lindblad equation, which can be written:

$$\dot{\rho} = -i[H, \rho] + \sum_{i,j} a_{ij}(F_i \rho F_j^\dagger - \frac{1}{2}\{F_j^\dagger F_i, \rho\}) \tag{2}$$

$$= -i[H, \rho] + \mathcal{D}_\gamma(\rho), \tag{3}$$

where the F_i are bounded operators, $\{,\}$ is the anti-commutator, and the matrix a_{ij} is positive semi-definite (Breuer and Petruccione 2002: Eq. 3.63). (We write the equation in this particular form rather than the more customary diagonal form (Breuer and Petruccione 2002: Eq. 3.61) since the properties of the matrix a_{ij} are useful for the analysis below.)

The Lindblad equation is made up of a unitary part identical to the von Neumann equation and a non-unitary part. For an initial pure state the unitary part reduces to the Schrödinger equation (hence, the Lindblad equation may be regarded as more general). In physical models the non-unitary part of the dynamics, encoded in the super-operator $\mathcal{D}_\gamma(\rho)$, corresponds to a dissipator term that encodes the parameterised effects of decoherence, thermal damping and noise.

The non-unitary dynamics of a system having the Lindblad form is a sufficient (but not necessary) condition for probability to be conserved.[9] Furthermore, there are general results suggesting Lindblad-type equations lead to well-posed partial differential equations in physical contexts (Azouit et al. 2016). The connection between the form of the dynamical map and the modelling of autonomous systems and their environments runs even deeper than the conservation of probability and well-posed differential equations. To see this, consider a selection of important results from the theory of open quantum systems.

A Markov process is a generalisation of a deterministic process for which the probabilistic state of a given system at some time is wholly determined by the dynamics together with the probabilistic state at some other time. Markovian systems are thus *memoryless* in a specific, non-time-directed, sense. For any Markov process the dynamical maps uniquely map each state in its state space to another state in the same space and can be composed, that is:

$$\Lambda_{t+s}\rho = \Lambda_t\Lambda_s\rho \tag{4}$$

Recall from above that a density operator, ρ, is defined to be a positive semi-definite operator of unit trace. For Λ_t to map density operators to density operators it must be a *trace-preserving positive map*. Trace-preservation is a necessary and sufficient condition for a dynamical map to be probability conserving. A *completely positive trace-preserving map* is then such that $\Lambda_t \otimes I_n$ is also a positive map for all n, where I_n is the identity map on the state space of some arbitrary further system of dimension n.

Stinespring's dilation theorem (Stinespring 1955) shows that if the dynamics of a system $\mathcal{S}$ is given by a completely positive trace-preserving map, then corresponding to ρ, there is a pure state of a larger system $\mathcal{S} + \mathcal{E}$ with unitary dynamics from which we can derive the non-unitary dynamics of $\mathcal{S}$. Following Cuffaro and Hartmann (2024), we can interpret the theorem as implying that *if* the map Λ_t is completely positive (CP), then the *open-system model* of S is *embeddable* into a larger *closed-system model* of $\mathcal{S} + \mathcal{E}$. Thus, if an 'embeddable system' is defined to be an open-quantum system that can be embedded into a larger model of a closed quantum system, then Stinespring's dilation theorem implies that open-quantum systems with CP dynamics are embeddable.[10]

There is then a natural connection between open quantum systems and parameterised autonomy as follows. The semi-classical limit of an open-quantum-system model is a dissipative classical model as is indicated by the interpretation of the non-unitary part of the Lindblad equation as a dissipator. Most vividly, we can show that the Caldeira–Leggett master equation recovers the equations of motion for a damped Brownian particle when the Ehrenfest-type relations are used to derive the equations of motion for the first- and second-order moments (Breuer and Petruccione 2002: 175). As such, we typically think of an open quantum model as the quantum analogue of an open classical model.

We would therefore expect to be able to derive open quantum models from the quantisation of open classical models. However, there is a severe formal challenge to understanding open quantum models as the quantisation of classical open models. Even in the simplest case of an open classical model with a linear, velocity-dependent friction term the quantisation operation is not well-defined since such systems are not symplectic. That is, such models do not fit with the usual canonical approach to quantisation built upon the symplectic structure encoded in the Poisson bracket and Hamilton's equations. The response to this challenge follows one of two approaches:

> The first one employs new quantization schemes, while the second one (the so-called system–reservoir approach) treats the particle as part of a larger quantum system which is quantized according to the usual rules. Of course, only the second approach can be considered as fundamental. (Joos 2013: 80)

This final line is telling. In the system–reservoir approach one starts with a classical closed-system model which is such that the spatially contiguous environment (the reservoir) is given a fine-grained representation and we have absolutely autonomous and closed dynamics. Canonical quantisation is then applied to derive a full quantum description for the system and reservoir. Through a combination of limits and tracing out, a model is constructed for the system degrees of freedom that encodes the quantum effects of friction but does not represent the reservoir in the quantum dynamics. Thus we derive an open system (conservation fails) which has *parameterised* autonomy. The most famous example of such a procedure is the derivation of the Caldeira–Leggett model, mentioned above, where the system–reservoir approach is deployed with the environment explicitly modelled as a set of non-interacting harmonic oscillators that are linearly coupled to the system (Breuer and Petruccione 2002, Joos 2013).

This is to follow the logic of Stinespring's dilation theorem the other way round. Rather than starting with an open-system quantum master equation that is shown to be embeddable in a larger closed-system dynamics, we first establish the larger closed-system dynamics based upon the classical model, and then construct the reduced open-system quantum dynamics by reducing a closed-system quantum model. Is there an alternative? Must all open systems—both classical and quantum—be understood as embeddable and, moreover, autonomous only in the parameterised sense? The final section considers these questions in the context of cosmology.

10.6 The Universe as an Open Quantum System

The different senses of openness discussed in Section 10.2 leave open the question of whether the universe is a closed system, because the way we answer the question depends on how 'the universe' and 'closed' are understood. If the universe is taken to be everything that there is, then it is not interacting with or affected by anything else. Thus, it is closed in the material sense of 1. above. However, that does not tell us anything about whether it is closed in any other sense, unless it is assumed that all that there is conserves every physical property (so that the universe is closed in every sense in the material mode, and should be modelled as closed in every sense in the formal mode 1.*).

In the context of quantum cosmology, that is, the application of quantum theory to the universe, it is standardly assumed that the universe must be modelled as evolving unitarily, because there is no other system for it to interact with, and isolated quantum systems are assumed to be subject to unitary time evolution. Thus, isolation is taken to imply 'closed' in all senses and conservation of all the relevant characteristic quantities and unitary time evolution, and the universe cannot be quantum-open.

We offer the following response to this argument. First, as the system of study in cosmology, the universe is not in fact all there is. Even in the material mode, the universe in the context of cosmology is not everything there is but only the large-scale physical structure of the universe. Second, as discussed in the previous section, quantum-mechanical systems that are interacting with an environment can also be modelled with unitary time evolution, thus the relationship between isolation and unitarity is complicated. As discussed in the previous section, in quantum mechanics 'closed' is a term of art that means 'subject to unitary time evolution', and it does not mean what is meant by 'closed' when it is said that the universe as a whole is closed because it is all there is. If what we mean when we ask if the universe is a closed quantum system is whether quantum-cosmological models are required to implement unitary time evolution, then the question turns out to be open for two reasons. The universe might be quantum-open since we might doubt that the model should be describing an isolated system since we may wish to include endogenous interactions and these may (or may not) require non-unitary time evolution. Then, even neglecting the question of non-isolation and endogenous interactions, the universe might be quantum-open since there is no formal or physical necessity connecting absence of interactions to unitary evolution. Indeed, models exist of classical cosmology in which measure compression, a close cousin of non-unitarity, obtains (Sloan 2018, Sloan 2023, Bravetti et al. 2023). Hence, whether the universe is quantum-closed is an open question.

The discussion of the previous section allows us to frame two more precise arguments that the universe must be modelled as quantum-closed. First, if open quantum systems are always embedded in larger closed quantum systems then the universe must be a closed quantum system, since there is nothing to embed it within. Second, if the derivation of an open quantum model is, fundamentally speaking, always via the system–reservoir approach then such models cannot describe the universe as the system in question, since there is no exogenous system to act as a reservoir. We conclude by rebutting each of these arguments.

We start with the embeddability argument. The most straightforward response is just to note that embeddability does not imply embeddedness. That is, although the fact that the dynamical map is completely positive may imply that formally we can embed an open-quantum-system model in a larger closed-quantum-system model, it may still be the case that no such larger system exists. A further response is that it remains to be seen whether a quantum model of the universe will correspond to a master equation with completely positive dynamics.[11]

There are physically important master equations like the Caldeira–Leggett equation that are not completely positive. We can see this straightforwardly by manipulating the equation as far as we can towards the Lindblad form (2) and finding that the matrix a_{ij} cannot be made positive semi-definite without the addition of extra terms (see Breuer and Petruccione 2002: 172–3, and Ferialdi 2017). Thus, whilst this particular equation was derived via the system–reservoir approach, it

provides a proof of principle that not completely positive master equations can be used to model real physical systems.

Furthermore, the standard argument that we should treat complete positivity as a physically motivated necessary condition on consistent master equations comes from the assumption that quantum dynamics must always be understood as fundamentally unitary. In general terms, there are reasons to doubt this assumption coming from what Cuffaro and Hartmann (2024) call the open systems view. Moreover, in the cosmological context, whether the universe is an open or closed system was the question we started with—thus complete positivity and embeddability does not give us an independent reason to expect the universe to be a closed system, but rather a more sophisticated way of encoding such a closed-systems perspective.

Let us turn to the quantisation argument. The most obvious response is to simply doubt the fundamentality assumption. Since there are alternative quantisation schemes for classical open systems that do not proceed via the system–reservoir approach (Dekker 1981), it may just prove to be the case that they are more fundamental, in the cosmological context. Moreover, as argued by (Bondar et al. 2012, Bondar et al. 2016) there is also a methodological justification for quantisations that proceed via the more direct route. In particular, the idea is to derive master equations based upon observed data recast in the form of Ehrenfest relations together with a specified mathematical structure of the equations of motion. The virtue of this approach, according to Bondar et al. (2016), is that it guarantees that the resulting equations of motion have the desired physical structure to reproduce the supplied dynamical observations (ibid., 1633). By contrast, the system–reservoir approach to the derivation of a master equation need not lead to a master equation that can reproduce the observations nor that the equations have a desired mathematical structure (ibid.).

The second response to the quantisation argument is that the universe need not be taken to be all that there is. We might understand the 'system' described by a cosmological model as being given by coarse-grained large-scale degrees of freedom and thus leave open the possibility of endogenous environmental interactions that lead to dissipative effects. Thus we might be able to apply the system–reservoir approach to derive an open quantum cosmology model, even when there is a more fundamental fine-scale dynamics that is unitary. More radically, we might instead insist that there simply is no fundamental fine-grained description of the universe and thus it is always necessary to apply some coarse-graining when trying to understand the universe as a whole.[12] Such a view has a nice historical precedent in terms of Ernst Mach's scepticism of universal applications of statistical reasoning (see Thébault 2021 and references therein) and is also in keeping with the idea of an 'never-ending tower' of effective field theories.[13]

Finally, and perhaps most excitingly, we can respond to the quantisation argument by considering the possibility, noted above, of classical open-system

cosmology models. In particular, classical cosmological models can be provided in the framework of contact geometry (Sloan 2018, Sloan 2023, Bravetti et al. 2023). Such models feature compression of the Liouville measure and they are thus open in the relevant classical sense. Moreover, they are absolutely autonomous in our terminology. The quantisation of such models, should it prove possible, should be expected to lead to models in which the universe is described as an open quantum system. Such models can also be expected to be absolutely autonomous and might be taken to provide a fundamental description of the universe in terms of non-unitary dynamics. Hence, the universe may be an open quantum system.

Acknowledgements

Thanks to Mike Cuffaro, Stephan Hartmann, Sébastien Rivat, and David Sloan for extensive discussion and to audience members in Bristol, Dubrovnik, Paris, and Santiago de Chile for helpful feedback. Thanks also to Anthony Thompson, Mike Cuffaro, and two anonymous referees for comments on a previous draft. Work on this paper was supported by the AHRC-DFG grant 'The Universe as an Open System' [Ref: AH/W010569/1(AHRC)/468374455 (DFG)].

Endnotes

1. The idea of material and formal modes is used by Ladyman and Ross (2007) following Carnap (1934).
2. This distinction is implicit in Cartwright's discussion of open quantum systems when she says, 'to demand a physical correlate of unitarity is to misunderstand what functions it serves in the quantum theory' (Cartwright 1983: 203). In our terms she is saying that the quantum open–closed distinction, which is just the unitary–non-unitary distinction (as Section 5 discusses in more detail), is made in formal mode and may not correspond to anything in material mode.
3. Ensuring that experiments were performed in a closed system in this sense, together with the precise measurement of masses and volumes, was critical to the Chemical Revolution, as was quantifying the heat flow associated with them (Ladyman et al. Forthcoming).
4. Not exchanging heat or work with the environment is sometimes described as 'closed', and sometimes described as 'isolated'.
5. This selection is closely related to Cartwright's idea of a nomological machine (Cartwright 1999).
6. Here we are following McNeill (2011).
7. This is Neurath's example discussed by Cartwright (1999). In this case the underspecified model fails to produce well-posed dynamical equations of motion, and so is not autonomous, but it does have representational capacity that tracks nomic features of its material mode counterpart, for example, time translation invariance.
8. If collapse is taken to be a real physical process then it may violate energy conservation (Carroll and Lodman 2021). We consider quantum mechanics without collapse.
9. See Cuffaro and Hartmann (2024) for discussion and a short proof.
10. It is worth considering in this context the view expressed by Cartwright (1983) that it is 'mistaken' to adopt an understanding of open quantum systems (such as that of Davies) that is based upon the assumption that open quantum systems can always be assumed to be part of a larger system

that is subject to unitary time evolution (ibid., 205). Clearly, Cartwright's judgement is consistent with the Stinespring theorem in so far as there are representations of open-system dynamics that can fail to be CP. For more details see Cuffaro and Myrvold (2013) and Cuffaro and Hartmann (2024).

11. There is a close formal connection between not completely positive master equations, dissipation, and non-Markovian dynamics (Ferialdi 2017). This means that the open-universe proposal we are considering may force us towards non-Markovian dynamics in quantum cosmology.
12. Thanks to Anthony Thompson for suggesting this point to us.
13. This idea is due to Cao and Schweber (1993) and has met with a degree of scepticism in the literature; see Hartmann (2001) for detailed discussion and Rivat and Grinbaum (2020) for an overview.

Bibliography

Alberti, A., Ivanov, V. V., Tino, G. M., and Ferrari, G. (2009), 'Engineering the Quantum Transport of Atomic Wavefunctions over Macroscopic Distances', in *Nature Physics* 5: 547–50.

Azouit, R., Sarlette, A., and Rouchon, P. (2016), 'Well-Posedness and Convergence of the Lindblad Master Equation for a Quantum Harmonic Oscillator with Multi-Photon Drive and Damping', in *ESAIM: Control, Optimisation and Calculus of Variations* 22: 1353–69.

Barbour, J. B. (2001), *The Discovery of Dynamics: A Study from a Machian Point of View of the Discovery and the Structure of Dynamical Theories* (Oxford University Press).

Batterman, R. W. (2018), 'Autonomy of Theories: An Explanatory Problem', in *Noûs* 52: 858–73.

Bondar, D. I., Cabrera, R., Campos, A., Mukamel, S., and Rabitz, H. A. (2016), 'Wigner–Lindblad Equations for Quantum Friction', in *The Journal of Physical Chemistry Letters* 7: 1632–7.

Bondar, D. I., Cabrera, R., Lompay, R. R., Ivanov, M. Y., and Rabitz, H. A. (2012), 'Operational Dynamic Modeling Transcending Quantum and Classical Mechanics', in *Physical Review Letters* 109: 190403.

Bravetti, A., Cruz, H., and Tapias, D. (2017), 'Contact Hamiltonian Mechanics', in *Annals of Physics* 376: 17–39.

Bravetti, A., de León, M., Marrero, J. C., and Padrón, E. (2020), 'Invariant Measures for Contact Hamiltonian Systems: Symplectic Sandwiches with Contact Bread', in *Journal of Physics A: Mathematical and Theoretical* 53: 455205.

Bravetti, A., Jackman, C., and Sloan, D. (2023), 'Scaling Symmetries, Contact Reduction and Poincaré's Dream', in *Journal of Physics A: Mathematical and Theoretical* 56: 435203.

Bravetti, A., and Tapias, D. (2015), 'Liouville's Theorem and the Canonical Measure for Nonconservative Systems from Contact Geometry', in *Journal of Physics A: Mathematical and Theoretical* 48: 245001.

Breuer, H.-P., and Petruccione, F. (2002), *The Theory of Open Quantum Systems* (Oxford University Press).

Cao, T. Y., and Schweber, S. S. (1993), 'The Conceptual Foundations and the Philosophical Aspects of Renormalization Theory', in *Synthese* 97: 33–108.

Carnap, R. (1934), *Logische Syntax der Sprache* (Springer).

Carroll, S. M., and Lodman, J. (2021), 'Energy Non-Conservation in Quantum Mechanics', in *Foundations of Physics* 51: 83.

Cartwright, N. (1983), *How the Laws of Physics Lie* (Oxford University Press).

Cartwright, N. (1999), *The Dappled World: A Study of the Boundaries of Science* (Cambridge University Press).
Cuffaro, M. E., and Hartmann, S. (2024), 'The Open Systems View', in *Philosophy of Physics* 2: 6:1–27.
Cuffaro, M. E., and Myrvold, W. C. (2013), 'On the Debate Concerning the Proper Characterization of Quantum Dynamical Evolution', in *Philosophy of Science* 80: 1125–36.
Darrigol, O. (2013), 'For a Philosophy of Hydrodynamics', in R. Batterman (ed.), *The Oxford Handbook of Philosophy of Physics* (Oxford University Press), 12–42.
Dekker, H. (1981), 'Classical and Quantum Mechanics of the Damped Harmonic Oscillator', in *Physics Reports* 80: 1–110.
Ferialdi, L. (2017), 'Dissipation in the Caldeira–Leggett Model', in *Physical Review A* 95: 052109.
Franklin, A. (2020), 'Whence the Effectiveness of Effective Field Theories?', in *The British Journal for the Philosophy of Science* 71: 1235–1259.
Hartmann, S. (2001), 'Effective Field Theories, Reductionism and Scientific Explanation', in *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 32: 267–304.
Jaeger, G. (2007), *Quantum Information: An Overview* (Springer).
Joos, E. (2013), 'Decoherence through Interaction with the Environment', in E. Joos, H. D. Zeh, C. Kiefer, D. J. Giulini, J. Kupsch, and I. O. Stamatescu (eds.), *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer Science & Business Media), 35–136.
Kapral, R. (2015), 'Quantum Dynamics in Open Quantum-Classical Systems', in *Journal of Physics: Condensed Matter* 27: 073201.
Ladyman, J., and Lorenzetti, L. (Forthcoming), 'Effective Ontic Structural Realism', in *The British Journal for the Philosophy of Science*.
Ladyman, J., Norman, N., and Seifert, V. (Forthcoming), 'A Philosophical History of Chemistry'.
Ladyman, J., and Ross, D. (2007), *Every Thing Must Go: Metaphysics Naturalized* (Oxford University Press).
Ladyman, J., and Wiesner, K. (2020), *What is a Complex System?* (Yale University Press).
McNeill, D. N. (2011), 'Antigone's Autonomy', in *Inquiry* 54: 411–41.
Newton, I. (1999), *The Principia* (University of California Press).
Nielsen, M. A., and Chuang, I. L. (2010), *Quantum Computation and Quantum Information* (Cambridge University Press).
Olby, R. (1971), 'Schrödinger's Problem: What Is Life?', in *Journal of the History of Biology* 4: 119–48.
Palacios, P. (2022), *Emergence and Reduction in Physics* (Cambridge University Press).
Phillips, A. C. (2013), *The Physics of Stars*, 2nd edition (John Wiley & Sons).
Preskill, J. (2018), *Quantum Information (Chapter 3)*. Lecture Notes for California Institute of Technology.
Reed, M. and Simon, B. (1975), *Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness*, Volume 2 (Elsevier).
Rivat, S. and Grinbaum, A. (2020), 'Philosophical Foundations of Effective Field Theories', in *The European Physical Journal A* 56: 1–10.
Robertson, K. (2020), 'Asymmetry, Abstraction, and Autonomy: Justifying Coarse-Graining in Statistical Mechanics', in *The British Journal for the Philosophy of Science* 71: 547–579.
Russell, B. (1903), *Principles of Mathematics* (Allen and Unwin).

Saari, D. G. (1973), 'Singularities and Collisions of Newtonian Gravitational Systems', in *Archive for Rational Mechanics and Analysis* 49: 311–20.
Sloan, D. (2018), 'Dynamical Similarity', in *Physical Review D* 97: 123541.
Sloan, D. (2023), 'Herglotz Action for Homogeneous Cosmologies', *Classical and Quantum Gravity* 40: 115008.
Stinespring, W. F. (1955), 'Positive Functions on C^*-Algebras', in *Proceedings of the American Mathematical Society* 6: 211–6.
Thébault, K. P. (2021), 'On Mach on Time', in *Studies in History and Philosophy of Science Part A* 89: 84–102.
Unruh, W. (2012), 'Decoherence Without Dissipation', in *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 370: 4454–9.
von Bertalanffy, L. (1950), 'The Theory of Open Systems in Physics and Biology', in *Science* 111: 23–9.

11

Biological Emergence: A Key Exemplar of the Open Systems View

George Ellis

11.1 Introduction

Because each organism is finite with a boundary, all biological entities are open systems (von Bertalanffy 1950, Peacocke 1989), with forces, matter, energy, and information impacting on them from the surrounding environment. They respond reciprocally by impacting on the environment in each of these ways. Because organisms carry out a variety of functions (Hartwell et al. 1999, Ball 2023), the Second Law of Thermodynamics implies they must use energy and materials in order to do so, obtained by interactions with the environment. These exchanges across the organism's boundaries are carefully controlled by physiological mechanisms at each emergent level. This chapter discusses how that happens.

Consequent on this openness, unpredictable external events impact on an organism, so its future behaviour cannot be fully determined by dynamical development from its initial internal state. To enable survival, true emergence must occur to cope with this situation, where (without making any metaphysical claims) by 'true emergence' I mean novel and robust behaviour at emergent levels. The organism must be capable of flexible adaptive responses to incoming influences, particularly by homeostasis and metabolic processes. Furthermore, higher-level organisms greatly enhance their survival prospects by developing memory-based agency (Humphrey 2022, Mitchell 2023, Corning et al. 2023), allowing predictive processing of incoming data regarding the changing external context, and so intelligent choice of actions.

The broad framework for this chapter is that set out in the paper 'Toward a Conceptual Framework for Biology' (Scheiner 2010) as follows (see Table 11.1). Item 1 is the central theme that I develop here. Items 2–8 are the keys allowing this to happen, with 9 and 10 providing context.

The aim of the chapter is to clarify the consequences of this openness of biological systems.

The following sections of this chapter look at, section 11.2: The Context: Modular Hierarchical Structures; which is what enables complexity to arise; section 11.3: The insufficiency of initial internal data to determine organismal

George Ellis, *Biological Emergence: A Key Exemplar of the Open Systems View*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press.
© Oxford University Press (2026). DOI: 10.1093/9780198929277.003.0012

Table 11.1 The domain and fundamental principles of the theory of biology (Scheiner 2010)

Domain
The diversity and complexity of living systems, including causes and consequences.
Principles
1. Life consists of open, non-equilibrium systems that are persistent.
2. The cell is the fundamental unit of life.
3. Life requires a system to store, use, and transmit information.
4. Living systems vary in their composition and structure at all levels.
5. Living systems consist of complex sets of interacting parts.
6. The complexity of living systems leads to emergent properties.
7. The complexity of living systems creates a role for contingency.
8. The persistence of living systems requires that they are capable of change over time.
9. Living systems come from other living systems.
10. Life originated from non-life.

outcomes, which is the key feature arising because of this openness; section 11.4: Processes to handle this issue functionally on short timescales; section 11.5: Processes to handle this issue on developmental and evolutionary timescales, which are much longer periods that set up the basis for the shorter ones; section 11.6: Symbolism and technology: the transforming factors at the highest levels, which is where the most sophisticated outcomes of openness occur; and section 11.7: The cosmological context, where the issue is: Should the universe itself be regarded as an open system over time?

11.2 The Context: Modular Hierarchical Structures

For good functional and evolutionary reasons, all truly complex systems, including life, are organised systems (Mossio 2023) that are adaptive modular hierarchical structures (Booch 1990, Simon 1996, Ellis 2016, Ellis and Di Sia 2023). The levels and details of the biological hierarchy are presented in depth in Campbell and Reece (2005). We can represent them, including the astronomical context, as in Table 11.2.

There are physical levels below level L1, but they do not matter as far as biology is concerned, because of the existence of classical and quantum protectorates (Laughlin and Pines 2000). Whatever their nature, which is only partially known, they lead to the emergence of physical levels L1–L2, as experimentally verified (Leighton and Sands 1965).

At each level below L14, one has spatially finite systems that interact with other systems at the same level (thus being open systems). The scale of the higher-level

Table 11.2 The hierarchical structure of biology in its astronomical context. Each higher level is the context for the next level down. Interlevel causation takes place as indicated by the arrows. The astronomical context is levels L12–L14. The arrows indicate the interlevel causation taking place.

	L14	Cosmos
	L13	Our Galaxy, Andromeda Galaxy
	L12	Solar system
	L11	Biosphere, International systems
↑	L10	Ecosystem, Nation
↑	L9	Population, Society
↑	L8	Organism, Person
↑	L7	Physiological systems
↑	L6	Tissues
↑	L5	Cells
↑	L4	Organelles, subcellular machinery
↑	L3	Macromolecules
↑	L2	Atoms
↑	L1	Particles (protons, neutrons, electrons)

open systems is larger than that of each of the lower-level ones. They emerge upwardly from lower-level systems as indicated by the upwards arrows, and influence lower levels as indicated by the downward arrows by either setting boundary conditions and constraints (Juarrero 2000, 2023, Noble 2002, 2008, 2012, 2016), or by creating, altering, or deleting lower-level elements (Ellis 2023). Thus each level is open relative to both the levels above them and below them, as well as to systems at the same level. Level L5 is the lowest one where all the functions of life occur (Mukherjee 2022): cells are open non-equilibrium systems (von Stockar 2013).

The biosphere itself (Level L11) is also open, driven by incoming high-grade solar radiation from the sky (L12 → L11), which provides its power source, and radiating low-grade radiation back to the sky (L11 → L12). Thus the sky acts as a heat sink for the solar system and for all biological systems on Earth. The low temperature of the sky enabling this is an outcome of the expansion of the universe.[1] This openness to the universe (L11 → L14) is thus key to the functioning of the sun and the biosphere in thermodynamic terms.

At a fundamental level, existence of all the emergent levels L2–L11 is due to the occurrence of broken symmetries (Anderson 1972, Zangwill 2021), which frees those higher levels from obeying the symmetry groups governing special relativity

and particle physics (Levels L1 and below). The boundaries defining an open system break these symmetries, as do their internal structures.

11.3 The Insufficiency of Initial Data to Determine Organismal Outcomes

The future behaviour of an organism cannot even in principle be uniquely determined by dynamical development from its internal state $S(t_0)$ at any initial time t_0, where $S(t_0) = \{x_I(t_0), p^I(t_0)\}$ specifies the position x_I and momentum p^I of every particle I in the organism's brain. This is the data required by Laplace's Demon to determine its future brain state from the relevant dynamical equations—Newton's law of motion, Maxwell's equations, the diffusion equation, and so on.[2] The essential point is that external events will occur that are not described by the data $S(t_0)$—a nearby car crash, tidal wave, tornado, the approach of a threatening dog or person, offer of a free holiday in Majorca, and so on. These require a suitable response—fight, flight, accept or reject an offer, and so on. The required brain micro-states $\{S(t) : t > t_0\}$ needed for appropriate responses to such external events at times $t > t_0$ are simply not available to the organism on the basis of $S(t_0)$ alone, because they omit all data about all such external influences that may impact the organism after the time t_0. These continually change the context of needed actions, and hence alter brain functioning.

True emergence (as defined above) must occur to cope with this situation. In order to react appropriately to such unpredictable incoming influences, as is needed to survive, the organism must be capable of adaptive predictive responses. Higher-level organisms therefore evolved sentience (Humphrey 2022) and agency (Mitchell 2023) to allow this kind of response, including the capacity of humans for predictive processing of likely outcomes (Hawkins and Blakeslee 2004, Clark 2013) and hence intelligent choice of actions. Furthermore, being an open system, any organism will from time to time experience hotter or colder conditions that must be adjusted to; the need for shelter if it rains or snows; dodging falling objects from trees or tidal waves from the sea; and so on. Thus all organisms must be able to adapt to such changes in their surroundings, which are not determined by the state of all the particles in their bodies at an initial time t_0. They must have agency-enabling adaptive responses to such events that affect them at later times. The Laplace's Demon argument fails to take into account spatial boundedness and openness. Indeed, the Laplacian conception of determinism is supposed to apply *only* to closed systems, i.e., if the system is closed (and the laws are deterministic), then we can predict its future from its initial state with certainty. So the Laplacian agrees with the contrapositive: if we can't predict the future state of the system from its initial state (and if the laws are deterministic), then the system is open.[3]

Furthermore, the total set of particles $P(t)$ that makes up the organism at time t will itself be different at a time $t_1 > t_0$ than at an initial time t_0 as some new particles will have been taken on board by breathing, drinking, and eating, and some will have been got rid of by breathing, sweating, and excreting. The set of particles $p(t_1)$ making up the organism at time t_1 is not the same as the set of particles $p(t_0)$ making up the organism at time t_0, and this later set of particles is not predictable from that initial state, which depends on the changing state of the environment (atmospheric pollution or eating at a restaurant, for example). Thus the context for Laplace's Demon for the organism as a whole to function is simply not there. The relevant particles are different, and the state space has changed. This is a central consequence of living systems being open systems. In Aristotelian terms, it is a form of time-dependent material causation (Ellis 2023).

Thus in the case of any open system, neither the information nor the material particles are available within the system for Laplace's Demon to begin to operate on and determine the future of that system from its initial state. Something else is going on.

11.4 Processes to Handle this Issue Functionally

To handle these problems of dealing with unknowable incoming data and material on functional (short) timescales, the organism needs metabolic systems and homeostasis. It will be much more successful if it has the capacity for predictive processing, and preferably agency, as in the case of fish, octopuses, lobsters, birds, and mammals including human beings—where planning and action selection can take place, and reciprocally shape the effects of the organism on the exterior environment, which in turn alters future environmental effects on the organism.

Metabolic systems handle taking in new material (solid, liquid, and gaseous) and transforming it into usable forms for functional purposes, moving them to where they are needed in the body, and then removing waste products and disposing of them. This happens in a coordinated way at each emergent level. Food ingested at the physiological level via the digestive system is made of biomolecules that are transformed to useful forms via enzymes at the cellular level, then transported via the cardiovascular system to where they are needed. The digestive system gets rid of waste products (urine and faeces). Because specific proteins are needed to make this happen, metabolic and gene regulatory networks work in an integrated way (Goelzer et al. 2008, Sauer and Teusink 2018). Underlying all this is the second law of thermodynamics: using energy for some purpose transforms useful stuff into unusable waste and generates entropy. Each cell is an open system that therefore needs nutrients to stay alive (von Stockar 2013, Mukherjee 2022), and uses energy in the form of ATP supplied by the citric acid

cycle (Akram 2014). Ingress to and egress from the cell is tightly controlled by voltage-gated and ligand-gated ion channels, which enable logical branching to emerge from the underlying physics (Ellis and Kopel 2019). This is a key link between physics and biology, enabled by cells being open systems. Metabolic processes occur via physiological systems such as mouth and nose and lungs and stomach and excretory systems at the macro-scale, together with metabolic cycles within cells at the micro-scale (Cabrera et al. 1998). The interaction with the environment is carefully controlled at each level in this way.

As far as the brain itself is concerned, there is a rapid turnover of brain molecules: indeed circuit-specific protein turnover is thought to underlie synaptic plasticity (Mohar et al. 2022). You simply are not dealing with the same molecules at later times than earlier times. However, the turnover takes place so that molecules are replaced one at a time without altering the state of the emergent structure (Crick 1984, Lee et al. 2022); in this way memory is preserved during these processes. Problems arise when this process starts to falter with aging (Rao et al. 2024).

Homeostasis is a mechanism for dealing with unpredicted incoming influences of an expected nature—they have been encountered before in our evolutionary history. It involves stabilising mechanisms enabling the system to respond appropriately to perturbations of a type often encountered by using an error signal (the difference between the current state and the desired state) to activate mechanisms that will change the situation back to the desired state (Cannon 1929, Modell et al. 2015). There are numerous feedback loops in the human body at both macro- and micro-levels, controlling blood pressure, temperature, heart rate, electrolyte levels, ion levels in cells, and so on (Guyton and Hall 2006, Hall and Hall 2020). You are ill if any of them are out of bounds.

Homeostatic systems can be adaptive in nature (Davies 2016). They can be taken to include the innate and adaptive immune systems (Parkin and Cohen 2001, Flajnik and Kasahara 2010), the latter dealing with the crucial issue of keeping healthy despite viruses of many unknown kinds invading the body unpredictably from the exterior. The immune systems are quintessential responses to the problem of the human body being an open system. Again outcomes are not predictable from prior knowledge of the state of all molecules in the body: the unpredictable incoming pathogens, as well as immunisation programmes instituted by health authorities to counter them, play a key role in shaping health and physiological outcomes at later times.

Agency allows a major advance in dealing with these issues, because it is the way we can handle completely unexpected events in a suitable way. Thus there has been strong evolutionary pressure for it to come into existence (Humphrey 2022, Mitchell 2023). This enables planning to meet new challenges from outside not merely when they are perceived but when they can be predicted, even if they

have never happened before. In humans it becomes effective in the context of a system-level architecture for cognition, emotion, and learning (Franklin et al. 2013, Franklin et al. 2016).

Information processing and the brain. Information processing is central to biology (Farnsworth et al. 2013, Nurse 2008), and of course in the functioning of the brain (Churchland and Sejnowski 1999), allowing higher levels of agency. Ongoing functioning of the brain at macro- and micro-scales enables appropriate responses to the ever-changing current situation in the physical and social environment. We interpret incoming information through predictive Bayesian brains (Clark 2013). Their state is updated on an ongoing basis as new data arrives via our various senses, filtered by cortico-thalamic feedback circuits (Alitto and Usrey 2003). Markov blankets act as an interface (Clark 2017), allowing minimisation of surprisal as we apprehend and then react to changes affecting us (Friston and Stephan 2007). The brain changes physically as this takes place at both macro- and micro-scales: this is brain plasticity (Kandel and Hawkins 1992, Kolb and Whishaw 1998), whereby we in a sense embody an image of the outside world in the details of our cortical connections. Memory of past events structures a mental model of the environment in which the organism is based, enabling predictions of future interactions and events (Hawkins et al. 2009). Off-line analysis on the basis of internal models enables planning of future actions in an anticipatory way (Mitchell 2023). This leads to mental causation occurring (Murphy and Brown 2007, Robb and Heil 2021) and affecting the outside world, which cannot uniquely anticipate what that action will be. Unpredictability across the organismal boundary goes both ways.

A key aspect of this interaction is the development of a Theory of Mind by each of us (Frith and Frith 1999, 2005), enabling appropriate navigating of social contexts. This internal model will be changing on a minute-by-minute basis as we interact socially and perceive social clues (Frith 2007).

At the mental level, humans are capable of imagination (Warnock 1976) that totally transforms outcomes of this processing: what the mind comes up with can be unlike anything that has ever existed before (iPhones, the internet, etc.). In that sense, the mental world is not a closed system: it can expand to include anything that is possible, thus transforming technology and society.

Interlevel integration. These processes take place at each level in Table 11.2. They have to be integrated so that the higher-level needs are fulfilled via the lower-level functions, and the lower-level functions are enabled to function as needed via the higher-level context. This integration must take place on an ongoing basis. This adaptive process at the mental level involves downward selection of desired lower-level states on the basis of higher-level needs (Noble 2021, Noble and Noble 2021),

enabled by the huge stochasticity at the molecular and cellular level (Mitchell 2018, Ball 2022, Graham 2023) that allows such adaptive selection (Mitchell 2023). Causal closure only occurs when all interacting levels are taken into account (Ellis 2020).

11.5 Processes to Handle this Issue on Developmental and Evolutionary Timescales

The organism needs developmental systems to bring it into being from a single cell (Wolpert et al. 2002). These are provided by gene-regulatory networks interlocking with metabolic systems which are read in a contextual way (Oyama et al. 2003, Griffiths and Gray 2005): genes and environment interact in complex feedback loops during development. The process of natural selection (Darwin and Wallace 1958) involves genetic mutations through various causes and then selection of those organisms with relatively higher reproduction rates leading to more fit organisms, all mediated by the inherited information transmitted to our offspring through our genes (Mitchell 2018, 2023). We inherit also DNA methylation, cellular machinery, and culture, as well as genes. Furthermore organisms construct their own niches, introducing evolutionary feedback loops: as open systems, the ecological context affects the organism and vice versa in a reciprocal way. The time development of each is reliant on the other because of their interactions. Neither is determinate on its own.

Evolution and development are integrated with each other to form Evo-Devo processes (Carroll 2005, 2008), so they should properly be thought of as taking place together—although they are of course only possible through minute-by-minute functional processes as just discussed, so in fact all three occur together. But the key point is that it is the evolutionary processes that allow adaptation of the organism, more or less successfully, to the physical, biological, and ecological environment. This is a crucial case of downward causation (Campbell 1974, Murphy and Brown 2007) in response to the organism being an open system subject to influences from these environments, where by 'downward causation' I mean higher levels affecting conditions at lower levels in a reproducible way, this being demonstrated either counter-factually or experimentally.

All the levels in Table 11.2 must be simultaneously adapted in an integrated fashion as the animal interacts with its environment on developmental and functional timescales, and as the population interacts with its environment on ecological and evolutionary timescales. Downward causation takes place from context **L10** to **L9** and **L8**, and then chains right down to the molecular level of genes and proteins (Wagner 2014) and the underlying physical levels.

11.6 Symbolism and Technology: The Transforming Factors on the Highest Levels

Humans are a symbolic species (Deacon 1998). Individuals are open systems, interacting to create a society (Berger 1963). Our interactions with each other led to the development of language, which transformed our lives by enabling communication of ideas, plans, and feelings in sophisticated ways. Furthermore symbolism underlies the power of narratives to structure our understandings and decisions. There are five essential points here.

First, the causal power of social structures (Elder-Vass 2010) arises from our symbolic capacities (Deacon 1998). These structures impact on our brains in many ways (Berger 1963, Berger and Luckmann 1966), and we influence these structures—another aspect of our being open systems, but now involving symbolic aspects—which certainly have causal powers. If society has passed a law about when you can cross the street, you are constrained to wait at a red traffic light and cross when it is green—with minute-by-minute symbolic information changing your brain state and hence your actions. More complex is how the laws of chess shape movements of physical pieces on a chess board via a learning process leading to those rules being embodied in the details of neural networks in our brains.

Second, this interaction enables transforming technologies to occur. This introduces a quite new effect: the ability to completely transform the future over long timescales by invention of new technologies which change possibilities in a transformative way (Arthur 2009), being carefully designed to do so (Dieter and Schmidt 2021). Future technology is completely impossible to predict: it depends on creative insights concerning how causal effects can be combined in completely new ways (Cropley 2015). There is an element of randomness in how this comes about which we do not understand—how a flash of insight comes to us, which can then be developed into a technological breakthrough. But we do know there is a great deal of randomness in brain operations at both the cellular and molecular level (Mitchell 2018) that must somehow underlie openness for such insights.

Third, new transforming ideas spread rapidly through nations across the world, so although there are certainly local customs and worldviews, there is an underlying openness of mental boundaries whereby ideas such as radios, cell phones, CAT scanners, GPS systems and so on spread world-wide on very short timescales and thereafter change life forever—for better or worse. Nations—and the minds in them—are in this way open systems. Transport by trains, motor cars, ships, and aircraft extends this openness to physical artefacts and people so that nations are no longer isolated—leading to the huge complex of social problems as to when immigrants are welcomed and when they are kept out. The whole issue of open systems becomes a key topic at the social, economic, and political levels of society.

Fourth, this means that we are open systems in the time dimension as well: cultural and technological discoveries change future possibilities on timescales

that transcend those of individuals, families, and dynasties. We can read a book that changes our minds on the basis of what happened hundreds or thousands of years ago: time travel occurs from the minds of Plato, Aristotle, Shakespeare, Charles Dickens, and so on, affecting us today at a time t_1 after we have read the book, which we had not done at time $t_0 < t_1$.

Fifth, overall this results in the transforming power of stories (Gottschall 2012) and narratives (Johnson et al. 2020), which are a key way the mind understands complex situations and makes decisions. As individuals or groups, we are open to the influence of narratives that transform our understanding of the situation, and hence our actions. News and social media are a key part of the equation that completely transcends the physical levels at the bottom, but certainly changes physical outcomes. The ecology of ideas and narratives of any specific society is an open system.

11.7 The Cosmological Context: An Open System Over Time?

Our cosmological context—the expanding and evolving universe that leads to the existence of galaxies, stars, and planets (Peebles 2020)—enables the existence of life. This is of course based in the underlying physics (Strassler 2024). Much has been written on why the universe and this physics is of such a nature that life can exist, but I will not pursue that here (Ellis 2014). The issue I now turn to is, does Laplacian determinism return if we consider the universe as a whole, when all the influences impacting on local open systems are themselves determined uniquely by the initial data for the universe itself? Does that data uniquely determine everything that happens? Or should the universe be regarded as an open system over time, with outcomes occurring allowed by but not necessitated by that data?

I will not engage in an extensive detailed argument in this regard: for that see Ellis (2021). I will just make three comments that I believe are conclusive.

First, quantum theory only predicts probable outcomes, not specific ones. That is the core of its nature (Isham 2001, Zagoskin 2015, Baggott 2016). Initial data in the early universe—say at the Last Scattering Surface—therefore does not determine unique outcomes at later times. Contextual wavefunction collapse takes place and results in the specific outcomes that occur with probabilities according to Bohr's rule (Drossel and Ellis 2018). Quantum chemistry (Karplus 2014) is the key link between physics and biological outcomes—a form of probabilistic contextual wavefunction collapse.

Second, fluctuations at the molecular level in living systems wipe out any memory of previous states at this level on extremely short timescales (Graham 2023). Biology makes use of this molecular chaos to extract order at higher levels from disorder at lower levels via molecular machines (Hoffmann 2012, Noble 2021, Noble and Noble 2021, Ball 2023). The fact that we are evolved open systems with

agency enables us to steer our own fate to a greater or lesser degree, depending on our family, social, economic, and intellectual history (Mitchell 2023). Physics per se is only one of the factors determining outcomes. Causal closure only takes place when all interacting elements and levels at different timescales are taken into account (Ellis 2020). Human agency involving progressive discovery of social and technological options (Arthur 2009) and economic and political decisions determine what happens at human scales.

Third, the knock-down argument is this. Current outcomes such as the design of iPhones and Jumbo jet aircraft, all of Shakespeare's sonnets, Einstein's theory of General Relativity, intelligent argumentation about open systems (as in this book), the corruption of the Supreme Court of the United States, could not possibly be a unique outcome of plausible physical processes taking place in the early universe. These lead to modulated Gaussian fluctuations on the Last Scattering Surface through well-understood astrophysical processes (Peter and Uzan 2017, Peebles 2020). These fluctuations cannot be encoded versions of all those later outcomes, which *inter alia* contain logical arguments such as in this book and detailed designs for digital computers (Harris and Harris 2013). Study astrophysics and cosmology as long as you like, and you will find no trace of any astrophysical process that can possibly lead to such intelligent logical outcomes.

Intelligent outcomes occur because the cosmos creates conditions where necessary elements for life, planets, ecosystems, and intelligent beings come into existence by evolutionary and developmental processes on different timescales, able to pursue literature, arts, and science (Ginsburg and Jablonka 2019). Through them, new things come into existence that cannot be anticipated on the basis of the detailed initial state of the early universe—because they are not uniquely specified by them. The physical universe is therefore open in time.

Acknowledgements

I thank two referees for helpful comments.

Endnotes

1. The existence of a dark night sky is known as 'Olbers' Paradox': Bondi (1960), Harrison (1965, 1974).
2. Quantum dynamics will only indirectly influence what is going on, because there is no wavefunction for a cell or a brain (see Ellis 2024). Classical physics is sufficient for this discussion, even though quantum physics adds extra uncertainty to physical outcomes.
3. I thank a referee for this comment.

Bibliography

Akram, M. (2014), 'Citric Acid Cycle and Role of its Intermediates in Metabolism', in *Cell Biochemistry and Biophysics* 68: 475–8.

Alitto, H. J., and Usrey, W. M. (2003), 'Corticothalamic Feedback and Sensory Processing', in *Current Opinion in Neurobiology* 13: 440–5.

Anderson, P. W. (1972), 'More Is Different: Broken Symmetry and the Nature of the Hierarchical Structure of Science', in *Science* 177: 393–6.

Arthur, W. B. (2009), *The Nature of Technology: What It Is and How It Evolves* (Simon and Schuster).

Baggott, J. (2016), The Quantum Story: A History in 40 moments, 2nd Edition (Oxford: Oxford University Press).

Ball, P. (2023), *How Life Works* (Chicago: University of Chicago Press).

Berger, P. L. (1963), *Invitation to Sociology: A Humanistic Perspective* (Anchor Books).

Berger, P. L., and Luckmann, T. (1966), *The Social Construction of Reality: A Treatise in the Sociology of Knowledge* (Garden City, N.Y.: Doubleday).

Bondi, H. (1960), *Cosmology* (Cambridge: Cambridge University Press).

Booch, G. (1990), *Object Oriented Design with Applications* (Benjamin-Cummings Publishing Co., Inc.).

Cabrera, M. E., Saidel, G. M., and Kalhan, S. C. (1998), 'Modelling Metabolic Dynamics. From Cellular Processes to Organ and Whole Body Responses', in *Progress in Biophysics and Molecular Biology* 69: 539–57.

Campbell, D. (1974), '"Downward Causation" in Hierarchically Organised Biological Systems', in F. J. Ayala, and T. Dobhzansky (eds.), *Studies in the Philosophy of Biology: Reduction and Related Problems* (London: Macmillan), 179–86.

Campbell, N. A., and Reece, J. B. (2005), *Biology* (Delhi: Pearson Education India).

Cannon, W. B. (1929), 'Organisation for Physical Homeostasis', in *Physiological Reviews* IX: 399–431.

Carroll, S. B. (2005), *Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom* (New York: W. W. Norton & Company).

Carroll, S. B. (2008), 'Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution', in *Cell* 134: 25–36.

Churchland, P. S., and Sejnowski, T. J. (1999), *The Computational Brain* (Cambridge, Mass.: MIT Press).

Clark, A. (2013), 'Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science', in *Behavioral and Brain Sciences* 36: 181–204.

Clark, A. (2017), 'How to Knit Your Own Markov Blanket: Resisting the Second Law with Metamorphic Minds', in T. Metzinger, and W. Wiese (eds.), *Philosophy and Predictive Processing*: *3* (Frankfurt/Main: MIND Group).

Corning, P. A., Kauffman, S. A., Noble, D., Shapiro, J. A., & Vane-Wright, R. I. (eds.) (2023), *Evolution on Purpose: Teleonomy in Living Systems* (Cambridge Mass.: MIT Press).

Crick, F. (1984), 'Memory and Molecular Turnover', in *Nature* 312: 101.

Cropley, D. H. (2015), 'Creativity in Engineering', in G. E. Corazza, and S. Agnoli (eds.), Creativity in Engineering (London: Springer), 155–73.

Cuffaro, M. E., and Hartmann, S. (2024), 'The Open Systems View', *Philosophy of Physics* 2: 6.

Darwin, C. and Wallace, A. R. (1958), *Evolution by Natural Selection (Memorial Volume)* (Cambridge: Cambridge University Press).

Davies, K. J. (2016), 'Adaptive Homeostasis', in *Molecular Aspects of Medicine* 49: 1–7.

Deacon, T. (1998), *The Symbolic Species: The Co-Evolution of Language and the Brain* (W. W. Norton).

Dieter, G. E. and Schmidt, L. C. (2021), *Engineering Design* (McGraw Hill).

Drossel, B., and Ellis, G. (2018), 'Contextual Wavefunction Collapse: An Integrated Theory of Quantum Measurement', in *New Journal of Physics* 20: 113025.

Elder-Vass, D. (2010), *The Causal Power of Social Structures: Emergence, Structure and Agency* (Cambridge University Press).

Ellis, G. (2014), 'On the Philosophy of Cosmology', in *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 46(A): 5–23.

Ellis, G. (2016), *How Can Physics Underlie the Mind. Top-Down Causation in the Human Context* (Berlin and Heidelberg: Springer).

Ellis, G. (2020), 'The Causal Closure of Physics in Real World Contexts', in *Foundations of Physics* 50: 1057–97.

Ellis, G. (2021), 'Physics, Determinism, and the Brain', in J. Voosholz, and M. Gabriel (eds.), *Top-Down Causation and Emergence* (Cham: Springer), 157–214.

Ellis, G. (2023), 'Efficient, Formal, Material, and Final Causes in Biology and Technology', in *Entropy* 25: 1301.

Ellis, G. (2024), 'Quantum Physics, Digital Computers, and Life from a Holistic Perspective', in *Foundations of Physics* 54: 56.

Ellis, G., and Di Sia, P. (2023), 'Complexity Theory in Biology and Technology: Broken Symmetries and Emergence', in *Symmetry* 15: 1945.

Ellis, G., and Kopel, J. (2019), 'The Dynamical Emergence of Biology from Physics: Branching Causation via Biomolecules', *Frontiers in Physiology* 9: 1966.

Farnsworth, K. D., Nelson, J., and Gershenson, C. (2013), 'Living is Information Processing: From Molecules to Global Systems', in *Acta Biotheoretica* 61: 203–22.

Flajnik, M. F., and Kasahara, M. (2010), 'Origin and Evolution of the Adaptive Immune System: Genetic Events and Selective Pressures', in *Nature Reviews Genetics* 11: 47–59.

Franklin, S., Madl, T., D'Mello, S., and Snaider, J. (2013), 'LIDA: A Systems-Level Architecture for Cognition, Emotion, and Learning', in *IEEE Transactions on Autonomous Mental Development* 6: 19–41.

Franklin, S., Madl, T., Strain, S., Faghihi, U., Dong, D., Kugele, S., Snaider, J., Agrawal, P., and Chen, S. (2016), 'A LIDA Cognitive Model Tutorial', in *Biologically Inspired Cognitive Architectures* 16: 105–30.

Friston, K. J., and Stephan, K. E. (2007), 'Free-Energy and the Brain', in *Synthese* 159: 417–58.

Frith, C. (2007), *Making Up the Mind: How the Brain Creates our Mental World* (John Wiley & Sons).

Frith, C., and Frith, U. (1999), 'Interacting Minds—A Biological Basis', in *Science* 286: 1692–5.

Frith C., and Frith, U. (2005), 'Theory of Mind', in *Current Biology* 15: R644–5.

Ginsburg, S. and Jablonka, E. (2019), *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness* (Cambridge Mass: MIT Press).

Goelzer, A., Bekkal Brikci, F., Martin-Verstraete, I., Noirot, P., Bessières, P., Aymerich, S., and Fromion, V. (2008), 'Reconstruction and Analysis of the Genetic and Metabolic Regulatory Networks of the Central Metabolism of *Bacillus Subtilis*', in *BMC Syst. Biol.* 2: 20.

Gottschall, J. (2012), *The Storytelling Animal: How Stories Make Us Human* (Houghton Mifflin Harcourt).

Graham, L. (2023), *Molecular Storms: The Physics of Stars, Cells, and the Origin of Life* (Cham: Springer).

Griffiths, P. E., and Gray, R. D. (2005), 'Discussion: Three Ways to Misunderstand Developmental Systems Theory', in *Biology and Philosophy* 20: 417–25.
Guyton, A., and Hall, J. E. (2006), *Textbook of Medical Physiology*. 11th edition. (Philadelphia, Pa.: Elsevier Saunders)
Harris, D. M. and Harris S. L. (2013), *Digital Design and Computer Architecture* (Amsterdam: Elsevier).
Harrison, E. R. (1965), 'Olbers' Paradox and the Background Radiation Density in an Isotropic Homogeneous Universe', in *Monthly Notices of the Royal Astronomical Society* 131: 1–12.
Harrison, E. R. (1974), 'Why the Sky is Dark at Night', in *Physics Today* 27: 30–6.
Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W. (1999), 'From Molecular to Modular Cell Biology', in *Nature* 402: C47–52.
Hawkins, J., and Blakeslee, S. (2004), *On Intelligence* (New York, NY: Times Books).
Hawkins, J., George, D., and Niemasik, J. (2009), 'Sequence Memory for Prediction, Inference and Behaviour', in *Philosophical Transactions of the Royal Society B: Biological Sciences* 364: 1203–9.
Hoffmann, P. M. (2012), *Life's Ratchet: How Molecular Machines Extract Order from Chaos* (Basic Books).
Humphrey, N. (2022), *Sentience: The Invention of Consciousness* (Oxford: Oxford University Press)
Isham, C. J. (2001), *Lectures on Quantum Theory – Mathematical and Structural Foundations* (London: Allied Publishers).
Johnson, S., Bilovich, A., and Tuckett, D. (2020), 'Conviction Narrative Theory: A Theory of Choice under Radical Uncertainty', in *Behavioural and Brain Sciences* 2020: 1–47.
Juarrero, A. (2009), *Dynamics in Actions* (Cambridge, Mass.: MIT Press).
Juarrero, A. (2023), *Context Changes Everything: How Constraints Create Coherence* (Cambridge, Mass: MIT Press).
Kandel, E. R., and Hawkins, R. D. (1992), 'The Biological Basis of Learning and Individuality', in *Scientific American* 267: 78–87.
Karplus, M. (2014), 'Development of Multiscale Models for Complex Chemical Systems: From H+ H2 to Biomolecules', in *Angewandte Chemie International Edition* 53: 9992–10005.
Kolb, B., and Whishaw, I. Q. (1998), 'Brain Plasticity and Behaviour', in *Annual Review of Psychology* 49: 43–64.
Laughlin, R. B., and Pines, D. (2000), 'The Theory of Everything', in *Proceedings of the National Academy of Sciences* 97: 28–31.
Lee, J., Chen, X., and Nicoll, R. A. (2022), 'Synaptic Memory Survives Molecular Turnover', in *Proceedings of the National Academy of Sciences* 119: e2211572119.
Leighton, R. B., and Sands, M. (eds.) (1965), *The Feynman Lectures on Physics* (Boston, MA: Addison-Wesley).
Mitchell, K. J. (2018), *Innate: How the Wiring of our Brains Shapes Who We Are* (Princeton: Princeton University Press).
Mitchell, K. J. (2023), *Free Agents: How Evolution Gave us Free Will* (Oxford: Oxford University Press).
Modell, H. et al. (2015), 'A Physiologist's View of Homeostasis', in *Advances in Physiology Education* 39: 259–66.
Mohar, B.et al. (2022), 'Brain-Wide Measurement of Protein Turnover with High Spatial and Temporal Resolution'. URL (bioRxiv) https://doi.org/10.1101/2022.11.12.516226.
Mossio, M. (2023) (ed.), *Organisation in Biology* (Springer).

Mukherjee, S. (2023), *The Song of the Cell* (London: Vintage).
Murphy, N., and Brown, W. S. (2007), *Did My Neurons Make Me Do It? Philosophical and Neurobiological Perspectives on Moral Responsibility and Free Will* (Oxford: Oxford University Press).
Noble, D. (2002), 'Modelling the Heart – From Genes to Cells to the Whole Organ', in *Science* 295: 1678–82.
Noble, D. (2008), *The Music of Life: Biology Beyond Genes* (Oxford: Oxford University Press).
Noble, D. (2012), 'A Theory of Biological Relativity: No Privileged Level of Causation', in *Interface Focus* 2: 55–64.
Noble, D. (2016), *Dance to the Tune of Life: Biological Relativity* (Cambridge: Cambridge University Press).
Noble, D. (2021), 'The Role of Stochasticity in Biological Communication Processes', in *Progress in Biophysics and Molecular Biology* 162: 122–8.
Noble, R., and Noble, D. (2021), 'Can Reasons and Values Influence Action: How Might Intentional Agency Work Physiologically?', in *Journal for General Philosophy of Science* 52: 277–95.
Nurse, P. (2008), 'Life, Logic and Information', in *Nature* 454: 424–6.
Oyama, S., Gray, R. D., and Griffiths, P. E. (eds.) (2003), *Cycles of Contingency: Developmental Systems and Evolution* (MIT Press).
Parkin, J., & Cohen, B. (2001), 'An Overview of the Immune System', in *The Lancet* 357: 1777–89.
Peacocke, A. (1989), *The Physical Chemistry of Biological Organization* (Oxford: Clarendon Press).
Peter, P., and Uzan, J. P. (2009), *Primordial Cosmology* (Oxford: Oxford University Press).
Rao, N. R., Upadhyay, A., and Savas. J. N. (2024), 'Derailed Protein Turnover in the Aging Mammalian Brain', in *Molecular Systems Biology* 2024: 1–20.
Robb, D., and Heil, J. (2021), 'Mental Causation', in E. N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University), Spring 2021 edition.
Sauer, U., and Teusink, B. (2018), 'Regulatory and Metabolic Networks', in *Current Opinion in Systems Biology* 18: iv–v.
Scheiner, S. M. (2010), 'Toward a Conceptual Framework for Biology', in *The Quarterly Review of Biology* 85: 293–318.
Simon, H. A. (1996), *The Sciences of the Artificial* (Cambridge Mass.: MIT Press).
von Bertalanffy, L. (1950), 'The Theory of Open Systems in Physics and Biology', in *Science* 111: 23–9.
von Stockar, U. (2013), 'Live Cells as Open Non-Equilibrium Systems', in U. von Stockar, and L. U. M. van der Wielen (eds.), *Biothermodynamics: The Role of Thermodynamics in Biochemical Engineering* (New York: EPFL Press), 475–534.
Wagner, A. (2014), *Arrival of the Fittest: Solving Evolution's Greatest Puzzle* (London: Penguin).
Warnock, M. (1976), *Imagination* (Berkeley: University of California Press).
Wolpert, L., Beddington, R., Jessell, T., Lawrence, P., Meyerowitz, E., and Smith, J. (2002), *Principles of Development* (Oxford: Oxford University Press).
Zagoskin, A. (2015), *Quantum Mechanics: A Complete Introduction* (London: John Murray Learning).
Zangwil, A. (2021), *A Mind over Matter: Philip Anderson and the Physics of the Very Many* (Oxford: Oxford University Press).

III

THE PHYSICS AND METAPHYSICS OF WORLDS AND UNIVERSES

12

Open Systems as Metaphysically Fundamental: Some Questions

Jørn Kløvfjell Mjelva, Josh Quirke, and Alastair Wilson

12.1 Introduction

If open systems are fundamental, as Cuffaro and Hartmann (2024) have proposed, then what is the fate of the closed system that corresponds to the entire physical universe?

One option is that the physical universe exists but is non-fundamental. This view has been discussed in the metaphysics literature under the name *priority pluralism*; it is the converse of Jonathan Schaffer's *priority monism* (2009). Priority monism itself has often been (including by Schaffer) motivated through appeal to quantum theory; arguments from entanglement to monism/holism can be traced back at least to Teller (1986). Nevertheless, Cuffaro and Hartmann aim to draw support for the open systems view from quantum theory, directly contrary to these monist arguments.

After a review in section 12.2 of the different fundamentality claims at issue, we look in section 12.3 at how Cuffaro and Hartmann's proposal handles existing arguments from quantum theory to monism/holism. We then turn in section 12.4 to the positive case that Cuffaro and Hartmann offer, based on quantum theory, for the open systems view. We raise the objection that their argument involves moving from a premise about epistemic possibility to a conclusion about objective possibility.

Even if the idea of a real but derivative cosmos can be made coherent through a suitable treatment of entanglement, such a cosmos remains puzzling. In section 12.5 we suggest an alternative metaphysics for the open systems view which lacks a complete cosmos altogether. Metaphysical realists about the content of physical theories typically assume that there is such a thing as the totality of physical reality. We explore the potential consequences for the metaphysics of physics of dropping this assumption. Perhaps the open-system theorist who wants to commit to the fundamentality of open systems ought to deny that there is any such thing as the totality of physical reality: for any candidate for physical reality, we can always characterise a more expansive reality relative to which our initial candidate is a mere subsystem. This is reminiscent of the indefinite

Jørn Kløvfjell Mjelva, Josh Quirke, and Alastair Wilson, *Open Systems as Metaphysically Fundamental: Some Questions*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press. © Oxford University Press (2026). DOI: 10.1093/9780198929277.003.0013

extensibility view in the philosophy of mathematics, associated in particular with Michael Dummett (1963) and also—in dialetheic form—with Graham Priest (2013). Indefinite extensibility has largely been studied to date in the context of logical and metaphysical ontologies—as an account of sets, for instance, or as an account of nested alethic modalities. We instead explore the application of the notion of indefinite extensibility to physics. We conclude that, while interesting and worth pursuing further, the indefinite extensibility view of the physical universe remains a challenging project likely to incur controversial philosophical commitments.

12.2 Which Fundamentality?

In 'The Open Systems View' (Cuffaro and Hartmann 2024), the authors make a sustained case that the framework of theorising about physical systems interacting with their environment, rather than about closed systems including all relevant environment, should be taken as fundamental. The purpose of this section is to explain what this means and why it might be believed.

Cuffaro and Hartmann distinguish three different kinds of fundamentality: ontic fundamentality, epistemic fundamentality, and explanatory fundamentality. Our concern here will be with ontic and explanatory fundamentality. Given a broadly explanationist/abductivist approach to metaphysics, and given that fundamentality is typically regarded as closely related to explanatory reductions between theories, these two notions are closely intertwined. It is worth unpacking these connections a bit. For explanationists about metaphysics, the ontological claims made by metaphysics are justified, in general, by their role in our best overall explanatory account of the phenomena. We should recognise the existence of those things which do explanatory work in our best theories. This explanatory work may be causal or non-causal in character.

It is uncontroversial that fundamental-level entities/quantities stand in explanatory relations—not necessarily being causes and effects in any simple way, but nevertheless playing some functional role in explanations mediated by the true underlying physical theory of our universe. Although there is a debate over whether this kind of explanation qualifies as causal, we can set this aside here, since that debate primarily concerns the specificity of the relationship to time that is built into the concept of causation. The explanations involved are generally agreed to be causal in the more general sense of being mediated by some law of nature (see Wilson 2020 for further discussion of that criterion).

In addition, a popular conception of ontological fundamentality links it to metaphysical explanation: the facts at the fundamental level are the facts which metaphysically explain all other facts. As with causal explanations, the conception of interlevel explanations as explanatory tends to establish a connection between

ontological fundamentality and explanatory fundamentality. Insofar then as ontologically fundamental entities are those which appear in our most fundamental explanations, we should expect that explanatory fundamentality and ontological fundamentality will tend to align. Given the background methodological considerations favouring keeping ontological and explanatory questions as aligned, we will be working in this chapter with a unified notion of fundamentality which combines both ontological and explanatory elements. We will refer to the resultant notion of fundamentality as *metaphysical fundamentality*.

We are assuming, then, that ontological fundamentality and explanatory fundamentality will coincide in all cases of interest. We take it that this is in line with Cuffaro and Hartmann's general approach, since in their discussion of explanatory fundamentality they are mainly concerned to deny that any notion of explanatory fundamentality over and above ontological fundamentality has clear relevance to the present debate. On that point we agree, and so we will remain neutral on how the issues we discuss will play out if ontological dependence is conceived as not essentially connected to explanatory dependence—as it is for example in Cameron (2022).

Cuffaro and Hartmann do not talk, as is customary in contemporary discussions of fundamentality in physics, in terms of the relative fundamentality of different kinds of physical fact. Instead their main focus is a notion of fundamentality which attaches to theoretical frameworks: roughly, the idea is that one framework is more fundamental than another if and only if the structures essential to the first framework determine the structures essential to the second framework in all relevant models. Here is Cuffaro and Hartmann's official statement of the relevant part of their main thesis, as applied to fundamentality of objects:

> (OntFund-O) Let $\mathcal{O}_F$ and $\mathcal{O}_P$ be any two objects of a given theoretical framework, $\mathcal{F}$. $\mathcal{O}_F$ is *ontologically more fundamental* than $\mathcal{O}_P$ with respect to $\mathcal{F}$ iff whenever an instance of $\mathcal{O}_P$ appears in any model of any theory that can be formulated in $\mathcal{F}$, some instance of $\mathcal{O}_F$ can be understood to determine $\mathcal{O}_P$ in that model. Furthermore, $\mathcal{O}_F$ is *ontologically fundamental* in $\mathcal{F}$ iff nothing is more fundamental than $\mathcal{O}_F$ in $\mathcal{F}$. (Cuffaro and Hartmann 2024: 16)

And here is their view, as applied to fundamentality of theoretical frameworks:

> (OntFund-2) Let $\mathcal{F}_F$ and $\mathcal{F}_P$ be two theoretical frameworks, and let $\{\mathcal{O}_F\}$ and $\{\mathcal{O}_P\}$ be their corresponding fundamental objects (in the sense, say, of OntFund-O). Furthermore let $\mathcal{F}_F$ and $\mathcal{F}_P$ be motivated by two distinct metaphysical positions, $\mathcal{M}_F$ and $\mathcal{M}_P$, respectively, in regard to their little-o objects. $\mathcal{F}_F$ is *ontologically more fundamental* than $\mathcal{F}_P$ iff the way that the $\{\mathcal{O}_P\}$ actually represent their little-o objects corresponds not to $\mathcal{M}_P$ but to $\mathcal{M}_F$. Furthermore, $\mathcal{F}_F$ is *ontologically*

> *fundamental* iff there is no theoretical framework more fundamental than $\mathcal{F}_F$. (Cuffaro and Hartmann 2024: 21)

In these definitions, an object $\mathcal{O}$ is of a framework $\mathcal{F}$ iff matters 'are representable in terms of $\mathcal{O}$ in every model of every theory that can be formulated in $\mathcal{F}$'. These are the abstract concepts used by that framework to represent the physical systems in a given model (the 'little-o objects').

OntFund-2 says, then, that a framework is more fundamental than another if and only if the less fundamental framework represents physical reality in a way which 'corresponds to' the metaphysical picture motivating the more fundamental framework. It is a little difficult to decipher what is meant by this idea of a correspondence between a way some physical models represent the world and a metaphysical view of the world. However, it seems open to interpret this in a straightforwardly realist kind of way: the fundamental physical framework is the one which is formulated in terms of the fundamental metaphysics of the world.

A peculiarity of this definition is that it does not seem to distinguish theories in terms of fundamentality if neither of them corresponds to the fundamental metaphysics of the world. Perhaps Cuffaro and Hartmann intend all such theories to be equally non-fundamental. But it is natural to think that something more can be said here to enable relative judgements of fundamentality even amongst non-fundamental frameworks. That feature was had by their first-pass definition of fundamentality, OntFund-1:

> (OntFund-1) Let $\mathcal{F}_F$ and $\mathcal{F}_P$ be two theoretical frameworks, and let $\mathcal{O}_P$ be any one of the fundamental objects (in the sense of OntFund-O) of $\mathcal{F}_P$. $\mathcal{F}_F$ is ontologically more fundamental than $\mathcal{F}_P$ iff $\mathcal{F}_F$ always re-describes any instance of $\mathcal{O}_P$ as effectively determined given an instance of some other object, $\mathcal{O}_F$, that is more fundamental in $\mathcal{F}_F$. Furthermore, $\mathcal{F}_F$ is ontologically fundamental iff there is no theoretical framework more fundamental than $\mathcal{F}_F$. (Cuffaro and Hartmann 2024: 16)

OntFund-1 can give rise to multi-layer orderings, such that $\mathcal{F}_F$ (the most fundamental theory) is more fundamental than both $\mathcal{F}_P$ and $\mathcal{F}_Q$, with $\mathcal{F}_P$ being more fundamental than $\mathcal{F}_Q$. A toy example might be general relativity (GR) → special relativity (SR) → classical mechanics (CM). However, it seems that OntFund-2 can only give us a two-layer ordering, with GR → SR and GR → CM but not SR → CM. (The underlying problem seems to be that the notion of correspondence between a framework's way of representing and a metaphysical framework is not comparative, so does not form transitive chains.) It is not clear that this flattening effect is intentional: perhaps either a return to OntFund-1, or a move to a third

formulation which allows for compound orderings, would better suit Cuffaro and Hartmann's intentions.

For present purposes, we take Cuffaro and Hartmann's core idea to be that fundamentality of a framework is linked to how the success of a framework is explained. If one framework's application to the phenomena is only explainable in terms of another framework, then the latter framework is more fundamental than the former. This seems very similar to saying that a framework is more fundamental than another when there is an explanatory reduction of the latter to the former, that is, when the former framework furnishes all the needed explanatory resources for an explanation of the core features of the latter framework. This sort of move from reduction to relative fundamentality is familiar from the effective scientific realist tradition (see, e.g., Robertson and Wilson 2026).

Recall that Cuffaro and Hartmann distinguish the kind of fundamentality they have in mind from the sense in which fundamental particles might be fundamental: 'The question of whether one theory or theoretical framework is more fundamental than another is not like the question of whether quarks are more fundamental than protons in particle physics'. However, we would still expect there to be a close connection between the fundamentality of a framework and the fundamentality of its little-o objects: if the little-o objects featuring in OntFund-2 are non-fundamental, then how these non-fundamental little-o objects are 'actually represented' by the fundamental objects of a given framework would offer little guidance as to the fundamentality of the framework. For instance, if only the universe as a whole is fundamental, then how a given framework 'actually represents' the non-fundamental subsystems of the universe has little bearing on the fundamentality of that framework.

Accordingly, we will proceed on the assumption that the fundamentality of a theoretical framework relative to another framework typically goes hand-in-hand with the fundamentality of the first framework's characteristic entities relative to the second framework's characteristic entities (staying neutral on what the explanation for that correlation might be). This will allow us to sidestep much of the complication around the definition of a framework and its relation to a set of models, and focus directly on the claim that open systems themselves—those real things out there—are more metaphysically fundamental than closed systems.

In the next two sections we turn to one particular physical system—the universe—and look at how extant metaphysical approaches classify it with respect to metaphysical fundamentality. We will first explore how Cuffaro and Hartmann resist arguments from quantum theory to monism, before turning to their own argument from quantum theory to pluralism, which we call the *open question argument*. We will then show how priority monists can respond to the open question argument, before exploring the metaphysical implications of priority pluralism in a bit more detail.

12.3 From Quantum Theory to Priority Monism

Quantum theory is a key battleground for the debate between priority monism and priority pluralism. Schaffer and other 'quantum holists' take quantum theory to support monism, whereas Cuffaro and Hartmann take it instead to support pluralism. This section and the next will attempt to get to the bottom of this situation, by identifying the different background assumptions which enable these authors to draw support from quantum theory for their preferred view.

Bringing fundamentality into the metaphysics of totality is a move with ancient roots, although it has recently revived in popularity—partly thanks to Jonathan Schaffer (2010, 2013), who defends a monism characterised in terms of fundamentality. Many monists—including the most influential monists of the 20th century—have denied the existence of anything except the universe—a view which Schaffer calls 'existence monism'. Schaffer endorses instead:

> *Priority monism*: The universe is fundamental, and parts of the universe are less fundamental.

Cuffaro and Hartmann, on our reading, cannot avoid denying priority monism. The open systems view entails that if there is a closed system corresponding to the totality of everything then that system is non-fundamental. And the totality of everything—call it the cosmos for short—is invariably conceived in the literature (including in Schaffer's discussion) as being a closed system. But perhaps this argument might be blocked, if the totality of physical existence itself were to be regarded as an open system. At times this option is suggested by Cuffaro and Hartmann's discussion. But what would it mean for the cosmos to be an open system, despite having no environment? One answer is simply that it can be correctly described by a non-unitarily evolving state vector. However, it is not clear to us that this technical sense of 'open' should trump the traditional criterion for a system being open, as stated in terms of systems and subsystems. According to that traditional criterion, a complete cosmos is a paradigm closed system.

Cuffaro and Hartmann recognise this difficulty, but deflect it using the idea that 'the ontological distinction between open and closed systems (at least when it pertains to the cosmos) breaks down in GT' (op. cit.: 19). They remark in this connection that metaphysical categories should not be imposed a priori upon physics, a claim with which we agree. Still, our point remains. Insofar as Cuffaro and Hartmann's proposal can be conceptualised in terms of existing categories (such that open systems but not closed systems have environments), it disfavours priority monism and favours:

> *Priority pluralism*: Parts of the universe are more fundamental than the universe itself.

Priority pluralism is a very respectable view within contemporary metaphysics—indeed, it is probably the orthodox view amongst those who

accept the notions of fundamentality used to frame the debate. Cuffaro and Hartmann position the metaphysical view behind the open systems framework as a departure from orthodoxy, and within philosophy of physics that may well be the case—but it is interesting to note that in general metaphysics there is already wide support for a position very like the one they endorse, albeit motivated by very different (usually a priori) considerations. Priority pluralism is for example endorsed, implicitly or explicitly, by Lewis (1986), Armstrong (1989, 1997), and Sider (2011). Philosophers in this camp tend to make an a priori case for metaphysical principles concerning the modal freedom of the fundamental entities—typically principles of recombination which permit any configuration of the fundamental entities within certain bounds—and the a priori case made for these principles tends to presuppose priority pluralism.

To the extent that philosophers of physics over recent decades have engaged with the monism–pluralism debate, the tendency has been to take physics—especially quantum physics—to support monism. For example, Howard (1989) and Schaffer (2010) argue directly from quantum theory to forms of monism (though Howard talks of 'ontological holism' rather than 'monism', his arguments can also be taken to support a monist view; cf. Ney, 2021), while Schaffer (2013) widens the scope of the argument to the claim that the underlying structure of both classical and quantum physics supports his priority monism. Cuffaro and Hartmann are going against the grain by defending a view more in the spirit of priority pluralism. As well as making a positive case for the open systems view, they accordingly owe us a response to these familiar pro-monist arguments which tend to favour the closed systems view.

The argument from entanglement, at its core, is that the intrinsic state of a whole entangled system contains all the information necessary to determine the intrinsic states of the parts—the entangled subsystems—but not vice versa. In Ismael and Schaffer (2020), the argument from entanglement is put in the form of an inference to the best explanation. The best explanation of the correlation between the individual subsystems of an entangled system is that they have a common ground: there is one underlying state, the state of the composite system, which both explains the states of the individual subsystems and their modal connection. This common ground is taken to be more fundamental than the entities it grounds. Moreover, Ismael and Schaffer argue such a common ground explanation would not face the same difficulties with Bell's Theorem as are faced by attempts at a common cause explanation. As far as we can see, Cuffaro and Hartmann have not offered a response to this style of argument from entanglement, and from our perspective it remains an important consideration which favours the closed systems view.

The argument from entanglement is not the only physics-based argument for priority monism that has been made out in recent years. In what he calls the argument from nomic integrity, Schaffer (2013) argues, drawing on Bigelow, Ellis, and Lierse (1992) and Maudlin (2007), that the fundamental laws of

physics—in both classical and quantum settings—apply directly to the totality of everything and only indirectly to subsystems of the universe. Bigelow, Ellis, and Lierse mention relativity principles, conservation laws, and symmetry principles as principles of physics which are fundamental and also global—in the sense that these principles 'neither ascribe properties to things within the world, nor describe correlations between things in the world' (384). Of these three kinds of principle, Schaffer focuses on conservation laws. Conservation laws in general only hold true at the global level: for any given open system, there is always the possibility of interaction with the environment which could lead to a change in the level of some (putatively) conserved quantity. Only considering the physical universe as a whole can be guaranteed to avoid this prospect completely. Therefore conservation laws apply to the universe as a whole. If any conservation laws are fundamental, then the system to which they apply—the whole universe—is equally fundamental. Or so Schaffer reasons. (Note that this argument applies equally to quantum and classical physics.)

We think that the most natural response for Cuffaro and Hartmann to Schaffer's argument from conservation laws is to deny that conservation laws are fundamental. Here Cuffaro and Hartmann may point to Noether's theorems, which on their most widespread physical interpretation (Lange 2007 provides evidence) are taken to show that conservation laws are non-fundamental in the sense that the holding of a conservation principle is explained by the holding of the corresponding symmetry principle. It is *because* of the symmetry of physical interactions under rotations that the quantity of angular momentum is globally preserved. So, on the assumption that physical laws which are wholly explained by other physical principles are non-fundamental, we get the conclusion that conservation laws are non-fundamental. Does this response simply move the bump in the carpet, from the conservation laws to the symmetry principles? If symmetry principles are essentially global, then we still have prospects for an argument from the fundamentality of symmetry principles to the fundamentality of the universe as a whole. Bigelow, Ellis, and Lierse do claim that symmetry principles are global, in the sense of applying to everything. Relatedly, the kind of symmetries in question (those to which Noether's theorems apply) are in physics usually called global symmetries, in the sense that they act the same way at every point. However, none of this makes the symmetry laws global in the sense relevant to Schaffer's argument—which is that they can only be understood as functions acting on the universe as a whole. And in a clear sense global symmetries do not apply essentially to the whole in this way: they apply individually at points (albeit in a globally uniform way) and the interactions which exhibit the relevant symmetries may themselves be locally defined.

This is all quite consonant with the open systems view so far as we can see. Entanglement is a threat to the locality of these interactions, of course; but then we are back to the argument from entanglement, and we have not yet identified a distinct argument from nomic integrity. Accordingly we would recommend that

defenders of the open systems view adopt an interpretation of Noether's theorems according to which conservation laws are non-fundamental and explained in some way by symmetry principles, where those symmetries are understood as local constraints on the behaviour of physical interactions, rather than being essentially global in character.

The upshot of this discussion is that Cuffaro and Hartmann have a good response available to the more general argument from nomic integrity, but there remains an apparent explanatory gap in their proposal with respect to the argument from entanglement, at least as posed by Ismael and Schaffer (2020). Of course, even if the open systems view does have an explanatory disadvantage over the closed systems view with respect to explaining the modal connection between subsystems of an entangled system, this disadvantage may be outweighed by other considerations. Accordingly we next turn to considering arguments in favour of the open systems view.

12.4 From Quantum Theory to the Open Systems View

We have discussed arguments from quantum theory to the closed systems view; now we turn to the potential for a converse case to be made, from quantum theory to the open systems view.

Cuffaro and Hartmann's main argument for the open systems view takes the form of an open question argument. First, they note that most (or even all) physical systems are represented as open systems even on the framework motivated by the closed systems view. As such, these physical systems could equally well be represented by the fundamental objects of the framework motivated by the open systems view. However, unlike the open systems view, the closed systems view is not able to make physical sense of dynamical evolutions of these (open) systems described by not completely positive maps (we grant Cuffaro and Hartmann this claim, though it is deniable). As it is currently an open theoretical question whether or not completely positive maps have essential application in physics, we should pre-emptively endorse the open systems view in order to properly allow for this possibility. Here is how they state the argument:

> … we think there is every reason to at least take the former possibility—the idea that the universe's evolution takes the form of the evolution of an open system—completely seriously in fundamental physics; i.e., as a live option that any theoretical framework should allow us to make sense of. And if we do this then there is every reason to embrace GT, rather than ST, as our preferred theoretical framework for quantum theory, given that the former is the only framework that actually permits us to model the dynamics of the universe fundamentally in these terms. (Cuffaro and Hartmann 2024: 20)

It is not only committed priority monists who should want to resist this type of inferential move within the metaphysics of physics. From our perspective, this argument for the open systems view amounts to a tempting but mistaken inference from open epistemic possibility to metaphysical possibility. If whether the open systems view is correct is a non-contingent matter (as Cuffaro and Hartmann appear to think) then that argument fails.

In general, we cannot infer from the fact that it is unknown whether p to the conclusion that p is a possibility in any non-epistemic sense. The open question argument for the open systems view may be compared to an open question argument for the truth of Goldbach's conjecture. (Assume realism about mathematical truth for the sake of the comparison.) It is currently an open question in mathematics whether Goldbach's conjecture is true or false. So we should take its truth as a live option. So we should adopt a framework for representing mathematical possibility according to which it is possible. But, since mathematical truths are non-contingent, if it is possible then it is true. So Goldbach's conjecture is true.

We take the lesson here to be simple: we cannot generate knowledge of genuine possibility out of ignorance, even principled ignorance. If the closed systems view is in fact correct, and non-contingently so, then the current live theoretical possibility that the open systems view is correct is now like the possibility Thales once entertained, that everything is made of water. We do not have to make room for the all-water scenario in our contemporary account of the different possibilities for matter just because that scenario was once a live possibility (see Wilson 2021 for a more extended version of this argument). The same, for all Cuffaro and Hartmann say, may go for the open systems view.

Perhaps this objection can be defused by regarding frameworks as not epistemically assessable—but instead as pragmatically adopted on grounds of fruitfulness. Still, the open epistemic possibility of fruitfulness does not entail the objective possibility of fruitfulness any more than open epistemic possibility of truth entails objective possibility of truth; and if questions about fruitfulness are not themselves epistemically assessable, we are dealing with a very radical kind of pragmatism.

Of course, none of this shows that the closed systems view is correct; in this section we have aimed only to undercut Cuffaro and Hartmann's positive argument for the open systems view.

12.5 Doing Without the Universe

Set aside, now, whatever reasons we might have for or against the open systems view, and explore its implications in a little more detail. If open systems are fundamental, what is the fate of the whole physical universe?

Cuffaro and Hartmann appreciate the pull of the idea that there is such a thing as the physical universe. Their official view leaves it open that closed systems, including the universe, might exist.

> As we have mentioned already, the open systems view does not deny that closed systems exist. It merely denies that only closed systems exist, and also that a closed system must exist, and represents all systems as, in general, open. (Cuffaro and Hartmann 2024: 14)

A question immediately arises about the modal status of the open systems view in light of these claims. If it is not the case that closed systems must exist, then is the cosmos a contingent existent? Or should we instead say that the cosmos is only contingently a closed system? But we set this question aside here. Our current question is about what it could mean for the totality of physical reality to be an open system. Cuffaro and Hartmann think that this latter question may have no good answer, because the concept of 'open system' might not apply in all cases:

> We must not presume that every such concept that arises in a scientific theory will conform, when interpreted ontologically, to our pre-theoretic concepts or even to the concepts of our predecessor theories. (Cuffaro and Hartmann 2024: 22)

Accordingly they suggest reframing the question in terms of representation:

> What, then, can it mean to represent the cosmos as a non-unitarily evolving density operator? (Cuffaro and Hartmann 2024: 22)

The options they consider are that we should imagine the non-unitary evolution of the universe construed as an open system as either resulting from non-physical external influence (perhaps from an interventionist deity) or as being a brute fact with no further explanation, at least none at the level of our current physics (Cuffaro and Hartmann 2024: 22). Neither of these options strike us—or indeed strike Cuffaro and Hartmann—as satisfactory. We would like to instead explore a different possible response, which is to deny that there is any such thing as the universe. Perhaps physical reality is *indefinitely extensible*, in the sense that any given physical system is a subsystem of a larger physical system.

The prospect of an indefinitely extensible material reality has been considered in contemporary metaphysics, but it is regarded as decidedly outré. Such hypothetical universes are sometimes referred to as *junky* (such that every object in them is a proper part of something—see Schaffer 2010), and they have attracted most attention in virtue of their role as a potential counterexample to mereological universalism. This attention has often been critical: Schaffer, for example, says that '[o]nly the most radical views of mereological composition, contravening both

intuition and science, could refuse the cosmos' (Schaffer 2013: 74). This dismissive attitude is supported elsewhere by considerations of the nature of a possible world: 'No world—provided that worlds are understood as possible concrete cosmoi—could contain worldless junk because a world that contained junk would be an entity not a proper part of another entity at that world. A world would top-off the junk' (Schaffer 2010: 65). Of course, one might think that worlds in the relevant sense are representations of concrete cosmoi rather than concrete cosmoi, and thereby avoid this objection. Be that as it may, it strikes us that the indefinite extensibility picture may be the most natural way to do justice to the intended metaphysical picture behind the open systems view, and in the space that remains to us we will explore this picture in a bit more detail.

A first point to note here is that indefinite extensibility of the cosmos does not necessarily mean indefinite extensibility in physical or temporal extent. Indeed an indefinitely extensible universe might in principle be finite in both space and time, so long as there were infinite complexity in the physical interactions within that finite spacetime. To take a toy model, consider a description in terms of non-relativistic n-particle quantum theory of an infinite number of non-interacting particles (imagine them to be bosons, such that spatiotemporal colocation is possible) moving freely in a finite closed spacetime. Such a universe might be indefinitely extensible in the sense that for any multi-particle composite system one can consider, there is still a further particle system with which the composite system could be combined. Since there would then be no totality of all the particles, nor would there be any physical system—open or closed—corresponding to the combined system of all the particles.

Can we simply collect together in thought all the systems in an infinitely extensible physical universe, and then introduce a term for the composite of them all taken together? According to the current proposal, such a procedure would not succeed in referring to any physical system, if indeed it succeeds in referring to anything at all. Since the collection cannot be picked out in any direct way—for example by enumerating a series of increasingly inclusive systems—we would have to introduce it by impredicative description. And, as is familiar from the study of such impredicative definitions, they may fail—most notoriously, in the paradoxical description 'the set of all sets'.

Of course, if it were impossible to refer to the most inclusive physical system, this impossibility would not be as not a result of any logical paradox, as it is in the case of set-theoretic indefinite extensibility. However, the impossibility of successful reference might—so far as we can see—simply be attributed to the nature of physical reality on the open systems view: it is not such as to form any totality, and hence attempts to refer to the cosmos as the collection of all physical systems taken together do not succeed. Either such attempts do not refer at all, or they succeed in referring to something which is not itself a physical system.

Indefinite extensibility for the physical universe naturally goes along with other forms of indefinite extensibility, for example with respect to the notion of

a set. Since it is plausible that for any physical object there is a set containing that object, then indefinite extensibility for physical reality seems directly to generate indefinite extensibility for set-theoretic reality. Likewise, since every physical system, no matter how large, will instantiate some new physical property (as represented by a density operator), the concept of physical property will also be indefinitely extensible.

An indefinitely extensible cosmos doesn't prevent us from doing cosmology as a science, but it may require some alterations in cosmologists' conception of their object of study. Cosmology becomes, on this picture, a matter of exploring physical reality on the very largest scales rather than a matter of exploring the universe as a whole. Thus cosmology becomes more open-ended: there could in principle be always more to say as we expand our horizons indefinitely. Yet cosmology does not have to be infinite in content under the open systems view: it could happen that from certain scales upwards there is nothing more interesting to say, in that certain patterns may repeat in the series of larger and larger systems one can consider. For example, a material universe isolated in an infinitely extensible empty spacetime would be one compatible with the open systems view, but one about which cosmology would eventually run out of interesting things to say.

12.6 Conclusion

The picture of physical reality as indefinitely extensible is extremely unfamiliar, and may strike many readers as bizarre. Nonetheless, we think it is a plausible way to make sense of the positive commitments of the open systems view, and to take at face value the idea that all physical systems are open. While we have questioned some of the positive arguments Cuffaro and Hartmann have marshalled in support of the open systems view, we hope to have highlighted how many interesting questions at the intersection of philosophy of physics and metaphysics arise from exploring its implications.

Acknowledgements

We are grateful to Michael Cuffaro and Stephan Hartmann for the opportunity to contribute to this volume and for encouraging and helpful editorial feedback throughout, and to two referees for their constructive comments. The chapter benefited from feedback from audiences at the Workshop on the Metaphysics of Quantities in Birmingham and the Annual Conference of the British Society for the Philosophy of Science in York, both in July 2024. We would in particular like to thank Sam Fletcher, Nick Huggett, Kerry McKenzie, Zee Perry, Jer Steeger, and Karim Thébault.

Bibliography

Armstrong, D. (1989), *A Combinatorial Theory of Possibility* (New York: Cambridge University Press).

Armstrong, D. (1997), *A World of States of Affairs* (New York: Cambridge University Press).

Bigelow, J., Ellis, B., and Lierse, C. (1992), 'The World as One of A Kind: Natural Necessity and Laws of Nature', in *The British Journal for the Philosophy of Science* 43: 371–88.

Cameron, R. (2022), *Chains of Being: Infinite Regress, Circularity, and Metaphysical Explanation* (Oxford: Oxford University Press).

Cuffaro, M. E., and Hartmann, S. (2024), 'The Open Systems View', in *Philosophy of Physics* 2: 6.

Dummett, M. (1963), 'The Philosophical Significance of Gödel's Theorem'. in *Ratio* 5 (2). Reprinted in *Truth and Other Enigmas* (1978) (Duckworth), 186–214.

Howard, D. (1989), 'Holism, Separability, and the Metaphysical Implications of the Bell Experiments', in J. Cushing and E. McMullin (eds.), *Philosophical Consequences of Quantum Theory* (South Bend: University of Notre Dame Press), 224–53.

Ismael, J., and Schaffer, J. (2020), 'Quantum Holism: Nonseparability as Common Ground', in *Synthese* 197: 4131–60.

Lange, M. (2007), 'Laws and Meta-Laws of Nature: Conservation Laws and Symmetries', in *Studies in History and Philosophy of Modern Physics* 38: 457–81.

Lewis, D. (1983), 'New Work for a Theory of Universals', in *Australasian Journal of Philosophy* 61: 343–77.

Priest, G. (2013), 'Indefinite Extensibility—Dialetheic Style', in *Studia Logica* 101: 1263–75.

Maudlin, T. (2007), *The Metaphysics Within Physics* (Oxford: Oxford University Press).

Ney, A. (2021), *The World in the Wave Function: A Metaphysics for Quantum Physics* (Oxford: Oxford University Press).

Robertson, K., and Wilson, A. (2026), 'Theoretical Relicts: Progress, Reduction, and Autonomy'. To appear in *The British Journal for the Philosophy of Science*.

Schaffer, J. (2010), 'Monism: The Priority of the Whole', in *Philosophical Review* 119: 31–76.

Schaffer, J. (2013), 'The Action of the Whole', in *Aristotelian Society Supplementary Volume* 87: 67–87.

Sider, T. (2011), *Writing the Book of the World* (Oxford: Oxford University Press).

Teller, P. (1986), 'Relational Holism and Quantum Mechanics', in *The British Journal for the Philosophy of Science* 37: 71–81.

Wilson, A. (2020), 'Classifying Dependencies', in D. Glick, G. Darby, and A. Marmodoro (eds.), *The Foundation of Reality: Fundamentality, Space, and Time* (Oxford: Oxford University Press), 46–68.

Wilson, A. (2021), 'Counterpossible Reasoning in Physics', in *Philosophy of Science* 88: 1113–24.

13
How Closed is Cosmology?

Sean Gryb and David Sloan

13.1 Introduction

What is a closed system? In trying to answer this question, we will find standard responses wanting in many respects. Our proposed solution—one that makes sense for self-gravitating systems—will have some seemingly unusual features. We will show that these features have unexpected virtues and only seem surprising because of the inadequacies of naive definitions of closed systems. In fact, our proposal will be seen to lead to some rather remarkable conclusions: the universe behaves like an open system where non-conservation of energy leads to the elimination of the conventional big bang and a solution to the problem of the arrow of time.

An important aspect of our proposal is the claim that energy need not be conserved in a closed system. Common intuition suggests that if a system is truly closed then there is nowhere for the energy to escape. But such an argument begs the question. Assuming that the energy, which is just some function on the state space of the system, must 'escape' somewhere when its value decreases is precisely the assumption that the energy must be conserved.

A more rigorous argument for conservation of energy in a closed system is to require the autonomy of the Lagrangian used to generate the system's laws. Closure of the system could be taken to motivate the formal requirement that the Lagrangian be a function only of an algebraically closed set of dependent variables (e.g., field configurations and their derivatives) with no explicit dependence on the independent variables (e.g., spacetime coordinates x or t) of the system. In this case, it's easy to show[1] that the equations of motion are equivalent to a symplectic Hamiltonian system where the Hamiltonian is a function only of phase space. A straightforward application of the theorems of symplectic geometry then suggests that the Hamiltonian function is a constant of motion. Identifying the Hamiltonian with the energy then gives conservation of energy.

This argument, though general and elegant, has several difficulties. First, it's not always appropriate to identify the Hamiltonian function with the energy of the system. Any reparametrisation-invariant theory, such as general relativity, will

Sean Gryb and David Sloan, *How Closed is Cosmology?*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press. © Oxford University Press (2026).
DOI: 10.1093/9780198929277.003.0014

have a vanishing Hamiltonian, and this cannot not lead to an interesting notion of energy. Second, the argument assumes that the laws of the theory are generated by Hamilton's principle, which under the assumption of the autonomy of the Lagrangian leads to a symplectic system. But many interesting physical systems, as we will argue later, are not symplectic. There is thus no guarantee, in general, that the Hamiltonian—let alone the energy—will be conserved in a non-symplectic system.

An alternative motivation for conservation of energy could run through Noether's first theorem and the assumption of time-translation invariance of the Lagrangian. Noether's first theorem, however, also assumes that the equations of motion are generated by Hamilton's principle and therefore also excludes non-symplectic systems. Moreover, it's not clear that time-translation invariance of the Lagrangian should be a necessary condition for a closed system. In order to have a definition of a closed system that is general enough to include non-symplectic systems, one must therefore abandon the notion that energy must be conserved.

There are, however, naive arguments against the existence of non-symplectic closed systems. These arguments run through Liouville's theorem, which gives the unique *Liouville measure* that is conserved in a (finite-dimensional) symplectic system.[2] The existence of such a measure guarantees that conserved probability distributions can be defined within the theory. Thus, within symplectic systems there is a straightforward way to 'conserve probabilities'. The conditions for Liouville's theorem are not, however, necessary conditions for being able to define conserved probabilities outside of symplectic systems. For example, in contact systems, which are not symplectic and will play an important role in our considerations below, generalisations of Liouville's theorem guarantee a unique dynamically conserved measure. Thus, there exists perfectly well-defined notions of probability for closed non-symplectic systems.

Our proposal for defining a closed system does not require the system to be symplectic. Instead, it makes use of an essential element of the proposals given above: the requirement of the autonomy of the dynamical system in question. While symplectic proposals rely on the autonomy of the Lagrangian and the use of Hamilton's principle to generate autonomous equations of motion, ours will focus more directly on the autonomy of the equations of motion themselves. In particular, we will define a closed dynamical system as a system whose dynamical laws are deterministic and algebraically closed in terms of a particular algebra of observables. We will be silent on the nature of the observable algebra in question so that the algebraic closure of our equations does not in principle prevent us from treating probabilistic or quantum theories (modulo the measurement problem).

To begin, we will give a simple example to illustrate different senses in which a dynamical system can be closed. This will help both to motivate our proposal and to illustrate the consequences for cosmology.

13.1.1 Closure in the Harmonic Oscillator

Let us first consider an example of a system that is generally accepted to be closed, and is so under all the definitions discussed above. This is the example of the harmonic oscillator consisting of two constants: the mass m and spring constant k; and an equation of motion relating acceleration a to position x,

$$a = -kx. \tag{1}$$

In this system, the mechanical energy,

$$E = \frac{mv^2}{2} + \frac{kx^2}{2}, \tag{2}$$

which is also numerically equal to the Hamiltonian of the system, is constant. If we plot the trajectory of such a particle on phase space (i.e., plotting momentum against position) we find that the orbits form ellipses, which we can directly see is the result of the energy, E, being constant along a classical trajectory. Given an ensemble of such systems, each with some initial position and momentum determined such that the ensemble spreads over some region R in the phase space, the dynamical flow of the ensemble is such that R is rotated about the origin at some fixed angular velocity while the volume it occupies on phase space is constant. This is a direct consequence of Liouville's theorem, which applies to this system.

Let us now imagine generalising the harmonic-oscillator model in the following standard way by adding a damping term proportional to the velocity v:

$$ma = -kx - fv \tag{3}$$

where f is a constant. There are now at least two senses in which the system is no longer closed. First, the mechanical energy of the system decreases over time because dE/dt is proportional to $-v^2 < 0$. Second, the orbits on phase space are no longer ellipses but spirals, which eventually come to rest at the origin. The region R then shrinks over time to a point at the origin.

But although our model is open in the sense that neither energy nor Liouville volume are preserved, it is nevertheless autonomous because the equations of motion are entirely self-contained: given an initial x and v and the value of f, we can solve the dynamical equations for all time without reference to any other information. This damped model is then autonomous, and therefore closed, in the sense we have defined above.

At this point, one might contend that damped oscillators should be understood as open systems because they are usually thought to interact dynamically with their environments, for example, through air resistance. But the model itself makes no

explicit reference to such effects, and including these explicitly would ultimately involve knowing more information about the environment and its interactions with the system. On the other hand, the damped oscillator can be taken as a complete system making no reference to the outside world, where the value of f is akin to a constant of nature. Absent further modelling assumptions, the system can thus be seen to be closed in the sense that all future behaviour can be inferred from current, self-contained knowledge.

The models treated in this section illustrate the three distinct notions of 'closure' discussed in the introduction. Note that these notions are completely independent: physical systems may exhibit any or none of these. But because the implications of each notion can differ drastically, it is important to identify which notions are relevant (or not) for the system in question.

13.2 Contact Systems

One class of systems with autonomous equations of motion that gives rise neither to conserved Liouville forms nor to conserved energies is the class of contact systems. Contact mechanics can be thought of as the odd-dimensional counterpart to symplectic mechanics. Here, for reasons of length, we will introduce only the necessary parts for our analysis, and work directly in Darboux coordinates. Comprehensive and general introductions can be found in Bravetti et al. (2017) and Gryb (2025).

Herglotz (Lazo et al. 2018) considered a generalisation of the usual Lagrangian mechanics to include Lagrangians that are dependent on the action, i.e.,

$$S = \int \mathcal{L}(q_i, \dot{q}_i, S)dt, \tag{4}$$

the minimisation of which gives rise to a slightly modified version of the usual Euler–Lagrange equations, henceforth called the Herglotz–Lagrange equations:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} - \frac{\partial \mathcal{L}}{\partial S}\frac{\partial \mathcal{L}}{\partial \dot{q}_i}. \tag{5}$$

Here we see that when the Lagrangian is independent of the action the usual equations are recovered. The Lagrangian is taken to be a function of the tangent bundle over configuration space and the action. This is an odd-dimensional space. By construction, our system is autonomous because the Lagrangian is a function only of the coordinates, their derivatives and the action. We can perform a Legendre transform and arrive at a contact Hamiltonian following the usual scheme. Let $p = \frac{\partial \mathcal{L}}{\partial \dot{q}}$. Then, the contact Hamiltonian is

$$\mathcal{H}^c = \sum_{i=1}^{n} p_i \dot{q}_i - \mathcal{L}. \tag{6}$$

From this we have the contact Hamilton equation:

$$\dot{q}_i = \frac{\partial \mathcal{H}^c}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial \mathcal{H}^c}{\partial q_i} - \frac{\partial \mathcal{H}^c}{\partial S}\frac{\partial H^c}{\partial p_i}, \qquad \dot{S} = \sum p_i \frac{\partial \mathcal{H}^c}{\partial p_i} - \mathcal{H}^c. \tag{7}$$

Again, the usual Hamilton equations are recovered when the contact Hamiltonian is independent of the action. It is a straightforward exercise to see that the contact Hamiltonian is not conserved under time evolution with

$$\dot{\mathcal{H}}^c = -\frac{\partial \mathcal{H}^c}{\partial S}\mathcal{H}^c. \tag{8}$$

The contact form, $\eta = \sum_{i=1}^{n} p_i dq_i - dS$, is the odd-dimensional counterpart to the symplectic potential. To obtain a volume form on our contact phase space, written in the coordinates q_i,p_i and S, we take

$$\Omega^c = \eta \wedge d\eta^n = -dS \wedge \bigwedge_{i=1}^{n} (dq_i \wedge dp_i). \tag{9}$$

Since the space is odd-dimensional, we cannot construct a Liouville-type measure because the symplectic form is a two-form, and hence any product of symplectic forms is an even-order form that cannot serve as a volume form for an odd-dimensional space. In contrast to the Liouville form, the privileged volume form on a contact manifold is *not* conserved under time evolution and obeys

$$\dot{\eta} = -\frac{\partial \mathcal{H}^c}{\partial S}\eta \;\rightarrow\; \dot{\Omega}^c = -(n+1)\frac{\partial \mathcal{H}^c}{\partial S}\Omega^c. \tag{10}$$

Let us illustrate our construction by considering the damped harmonic oscillator. If we begin with the Lagrangian

$$\mathcal{L} = \frac{m\dot{x}^2}{2} - \frac{kx^2}{2} - \frac{\mu S}{m} \tag{11}$$

then the Herglotz–Lagrange equation for x is precisely that of the damped harmonic oscillator (see equation 3). Identifying $p = m\dot{x}$, the contact Hamiltonian is thus

$$\mathcal{H}^c = \frac{p^2}{2m} + \frac{kx^2}{2} + \frac{\mu S}{m}. \tag{12}$$

It is important to note that this does not correspond to what we would think of as the mechanical energy of the oscillator as it contains a term proportional to the action. The evolution of the contact Hamiltonian shows that it has an exponential decay, as does the contact volume form.

We can interpret the decay of the contact volume form as a focusing of solutions on the contact space. This focusing of solutions, in turn, introduces a local time asymmetry into the system along particular solutions, and thus can be used to define a local arrow of time. More globally, because this focusing is monotonic

along *all* solutions and points in the same direction as the decrease in the contact Hamiltonian, the asymptotic future of any damped oscillator is a system at rest, $x = 0 = \dot{x}$. The future of all solutions converges on to this point. Hence, for all solutions we can further identify a global arrow of time: the direction pointing towards this final state. Equivalently, this arrow of time is the direction in which the energy (either mechanical, or the contact Hamiltonian itself) is decreasing. See Gryb (2025) for an extensive treatment of how an arrow of time can emerge through such a mechanism. The damped oscillator thus has a qualitative feature distinct from its undamped counterpart.

A given contact system can be symplectified. Doing so will produce a symplectic system in which the Hamiltonian is conserved under evolution. This is achieved at the cost of embedding the contact phase space into a phase space with one extra dimension. This process will imply that a single solution to the contact system is mapped onto multiple solutions of the symplectic system, which will appear as a symmetry in the symplectic representation. A simple way to achieve this is to expand our configuration space by introducing a new coordinate Q orthogonal to the contact space. We can then define the symplectic Hamiltonian and symplectic potential by

$$\mathcal{H}^s = Q\mathcal{H}^c, \qquad \theta = Q\eta. \tag{13}$$

It is a simple exercise to show that both the symplectic Hamiltonian and potential are conserved as a consequence of Hamilton's equations, and that the equation of motion for each of the contact variables remains unchanged by this process. In our damped harmonic oscillator example, our symplectic system is then

$$\mathcal{H}^s = \frac{\pi^2}{2mQ} + \frac{kQx^2}{2} + \frac{\mu QS}{m}, \qquad \omega = dS \wedge dQ + d\pi \wedge dx. \tag{14}$$

Note that the observable algebra of this system is not generated by the original x and $p = v$ variables but by x and $\pi = Qv$. This system, as any system made from such symplectifications, exhibits dynamical similarity of order 1 under the similarity vector $D = Q\frac{\partial}{\partial Q} + \pi\frac{\partial}{\partial \pi}$, i.e.,

$$\mathfrak{L}_D\mathcal{H}^s = \mathcal{H}^s, \qquad \mathfrak{L}_D\omega = \omega\,. \tag{15}$$

This reflects the redundancy brought to the system through the introduction of the coordinate Q.

At this point it is natural to ask why, if an open (contact) system can always be embedded within a closed (symplectic) system of this type, do we not simply declare the closed system to give a complete description? Our answer comes from demanding autonomy only on the smallest algebraic set required to retain empirical adequacy (Gryb & Sloan 2021). Let us motivate this demand now.

In our symplectification process we have, by necessity, introduced an auxiliary structure Q into the system. In the case of the damped harmonic oscillator, the momenta depend upon this auxiliary variable Q. Similarly, both the conserved

volume form and the conserved energy (i.e., the symplectic Hamiltonian) are also dependent upon Q, which can take arbitrary values. All these quantities will therefore take arbitrary values themselves, and thus cannot be used to represent physical structures. The value of Q can therefore have no empirical significance. We see that autonomy in the symplectic system is only achieved in terms of an algebra that is strictly larger than the minimum required for empirical adequacy. In terms of the measure, this means that states are counted as distinct even if they differ only by the (arbitrary) value of Q. This is clearly not a desirable state of affairs as it would introduce a distinction without a difference.

In the case of the damped oscillator, we thus find that the non-conservative, contact formulation should be treated as a complete description of the system as it involves an autonomous system that is independent of auxiliary structures. We conclude that certain apparently 'open' systems can reasonably be considered to be 'closed' in the sense that they provide complete descriptions.

13.3 Cosmology

Let us now apply our framework to cosmology. In our analysis we shall restrict to the simplest case: flat ($k = 0$) Friedmann–Lemaître–Robertson–Walker (FLRW) models. The results can be generalised to include open ($k = -1$) and closed ($k = 1$) spatial slices, and include anisotropies, together with general matter sources (Sloan 2021, Mercati 2019, Sloan 2023). Our considerations here thus illustrate a much more general procedure that can be applied to a variety of different situations. For reasons of space, we refer the interested reader to the cited papers for details.

There is an unfortunate clash of nomenclature here wherein cosmologically the terms 'open' and 'closed' (and separately 'flat') are used to refer to the geometry of a spatial manifold, which can either be a three-dimensional hyperboloid, a three-sphere or a compactified section of $\mathbb{R}^3$. The issue of whether or not a spatial slice is compact is not at issue here since the stringent demands of homogeneity and isotropy render such considerations orthogonal to the issue of 'open' versus 'closed' in terms of conservation laws or autonomy.

Cosmology provides an interesting test case for distinguishing between the various demands of being 'closed'. Since the object under consideration is taken to be the entire universe, it may be argued that if some quantity is to be conserved in a complete accounting of physics, then there should be nowhere to hide any losses. However, one may counter that cosmology is simply the consideration of long-wavelength modes of a full theory of general relativity, and that a complete model should account for the transfer of, e.g., energy between modes of differing wavelengths. Thus, a failing of strict conservation laws within the cosmological context does not rule out such conservation for the full theory.

Such an argument does not appear to hold in the vicinity of space-time singularities where, per the Belinski–Khalatnikov–Lifschitz (Belinsky et al. 1982, Berger et al. 1998, Ringstrom 2001, Heinzle & Uggla 2009, Uggla 2007) conjecture,

the dynamics of the full theory of general relativity asymptotes to those of a homogeneous solution at each spatial point. Further, since cosmological solutions are exact solutions of the full theory, as opposed to truncations in which some terms have been disregarded, any features of these solutions are necessarily features of at least a subset of solutions to the full theory. Thus we have a counter-example to the claim that conservation should hold in a full theory.

For our purposes it will suffice to demonstrate our results in the flat FLRW model with simple scalar fields as matter.[3] The behaviour of such a cosmology can be found from the Einstein–Hilbert action following a symmetry reduction to homogeneous and isotropic subsystems. After some technical considerations, the foremost of which is checking compatibility with the principle of symmetric criticality, we can render the action as

$$S_{\mathrm{EH}} = \int v\left(-\frac{\dot{v}^2}{3v^2} + \frac{\dot{\phi}_i^{\,2}}{2} - V\right) dt\,, \tag{16}$$

where $V = V(\vec{\phi})$ is a potential function and, to streamline our algebra, we adopt the convention $8\pi G = 1$. We here note that a cosmological constant can be included trivially in this system by having a non-zero minimum of V. The geometric sector is described by v, which represents the volume of a fiducial cell in comoving coordinates. Note that although the scale factor is often referred to as being the 'size of the universe', this is undefined in the case of an infinite spatial slice. From this action we can derive (following a Legendre transformation) the Hamiltonian description of our theory. We write this in terms of a Hamiltonian function and symplectic structure on phase space,

$$\begin{aligned} 0 := \mathcal{H} &= v\left(-\frac{3P_v^2}{8} + \frac{P_i^2}{2v^2} + V\right), \\ \omega &= dP_v \wedge dv + dP_i \wedge d\phi_i\,. \end{aligned} \tag{17}$$

Here, $P_v = -\frac{4\dot{v}}{3v} = 3H$, where H is the Hubble parameter. Note that the Hamiltonian is a constraint and must be zero—a fact that follows from the time reparametrisation invariance of general relativity. From the existence of these structures it is easy to verify that the 'energy' (the Hamiltonian) and the volume form $\Omega = \omega^{1+n}$ where n is the number of scalar fields, are conserved over time.[4] Furthermore, from Hamilton's equations we can find equations of motion for each of the phase-space variables written in terms of itself and the other phase-space variables. Our system is therefore autonomous. It would hence seem apparent that cosmology is a closed system following all three notions that we have discussed.

There is, however, a catch: The algebra generated by the set of variables just described is strictly *larger* than what is needed for empirical adequacy. In particular, the volume v can be set to any value by re-scaling its initial condition without affecting any empirical predictions of the theory. This is because such predictions only ever rely on the value of v relative to some reference value (usually taken to be an initial or final condition).

The equations of motion of our system are the Friedmann equations and the Klein–Gordon equation:

$$H^2 = \frac{1}{3}\left(\frac{\dot{\phi}_i^2}{2} + V\right), \tag{18}$$

$$\dot{H} + H^2 = -\frac{1}{3}\left(\dot{\phi}_i^2 + V\right), \tag{19}$$

$$\ddot{\phi}_i + 3H\dot{\phi}_i + \frac{\partial V}{\partial \phi_i} = 0. \tag{20}$$

It is clear from these that the dynamical algebra closes in terms of the ϕ and H without ever needing to refer to v. We should therefore seek a description of our system that never makes reference to v in the first place. To do so, we note that the freedom to choose the value of v at any time represents a scaling symmetry, and thus we can follow the analysis of Sloan (2018), Bravetti et al. (2023), and Gryb (2025) and reduce our symplectic system to a contact system with one fewer dimension. In fact, it was shown in Sloan (2021) and Sloan (2023) that the dynamics of this system can be derived from an action principle that never refers to scale in the first place by extremising the action

$$S = \int_0^t \left(\frac{3S_h^2(t')}{2} + \frac{\dot{\phi}_i^2}{2} - V\right) dt'. \tag{21}$$

This is an action of the Herglotz type as discussed above. Together with the identification $S = -H$ this produces the original equations of motion. The action is equivalent to the second Friedmann equation, and the Herglotz–Lagrange equation for ϕ is the Klein–Gordon equation. The contact Hamiltonian is then equivalent to the first Friedmann equation:

$$0 := \mathcal{H}^c = -\frac{3S^2}{2} + \frac{p_i^2}{2} + V, \qquad \eta = -dS + dp_i \wedge d\phi_i, \tag{22}$$

with $p_i = \dot{\phi}_i$.

There are several notable features of this description of our system. The contact Hamiltonian and contact structure are expressed directly in terms of the minimum autonomous system $(\phi, \dot{\phi}, S = -H)$ required for empirical adequacy rather than the usual symplectic variables used in cosmology. The identification of the action S with the Hubble parameter makes the nature of 'Hubble friction' more apparent. From equation 10, we see that the natural volume form over physical quantities will shrink in time. This solution focusing is a hallmark of 'open' systems.

The action is monotonically increasing, and can therefore act as a physically observable proxy for time. It is common when addressing issues of measure in cosmology to construct a space of solutions by restricting to a surface whereon the Hubble parameter is constant (Gibbons et al. 1987, Ashtekar & Sloan 2010). This is the restriction to a constant extrinsic curvature slicing in the cosmological context. Doing so demonstrates clearly the distinction between the symplectic

and contact cases: If we restrict the Liouville form in each case to this slice, we find that $\Omega^s = \dot{\phi} dv \wedge d\phi$ whereas $\Omega^c = \dot{\phi} d\phi$ since the contact phase space is one dimension smaller. Thus in the symplectic case, when evaluating measures on phase space, one must choose an interval in v over which to integrate. This can lead to problems because the volume is not empirically significant. However, since the amount by which volume changes between two times is, the freedom to choose different initial conditions for v leads to ambiguous statements (Ashtekar & Sloan 2011, Sloan & Silk 2016, Corichi & Karami 2011, Corichi & Sloan 2014). In the contact case, the measure is unambiguous but not preserved under time evolution. The time evolution of the measure is proportional to the action, and hence solutions focus on those which maximise the integral of the Hubble parameter between two time slices.

These considerations show that in cosmology, as in the damped oscillator, 'open' dynamical systems can reasonably be understood to give complete, though idealised, descriptions of the universe.

A further enticing feature of the open description of cosmology is that the system of differential equations it produces remains deterministic at, and beyond, the initial big bang singularity. The status of singularities within general relativity as a whole is a complex topic with the Hawking–Penrose theorems proving that, under fairly generic circumstances, there will exist points in a space-time beyond which the paths of physical objects within the system become non-deterministic as the geodesics that objects follow are incomplete. Thus, Einstein's equations do not admit a unique continuation beyond such points. In the cosmological setup, our considerations are somewhat simpler: we need only concern ourselves with the determinism of the system of equations derived from the symplectic Hamiltonian description (equations 17). The relevant point in the evolution is when the Hubble parameter becomes infinite, as does (necessarily) at least one contribution to the matter–energy density. A theorem due to Foster (1998) shows that under a broad range of conditions this corresponds to the kinetic energy of the scalar field becoming infinite.

Infinities, even of physically observable quantities, do not necessary signal the breakdown of the evolution equations. If we consider the harmonic oscillator described in terms of $y = 1/x$, we see that its evolution results regularly in infinities of y in finite time. The physical system, with x interpreted as the distance to equilibrium, sees such points as nothing special and continues beyond them. The infinities may appear, on first glance, to inhibit the integration of the equations of motion. However, when rendered in the usual manner through a change of variables, we see that they are in fact an artefact of the description.

In our cosmological model it is the scalar fields that become infinite. We can change coordinates through a gnomonic projection. For two scalar fields, this is achieved by projecting a point on a plane onto one half of a sphere. The unit radius sphere is tangential to the plane, touching at the origin. Each point P on the plane is then mapped uniquely to a point Q on the lower half of the sphere by considering a line from the centre of the sphere to the point P, and taking Q to be the point of

intersection of this line with the sphere. Thus two scalar field values, ϕ_1 and ϕ_2 are represented as a pair of longitude and latitude on the sphere. The singularity in the original coordinates (where at least one of the fields has become infinite) is mapped to the equator of the sphere. It is then a matter of technical calculations (Koslowski et al. 2018, Sloan 2019) to show that the equations of motion of the system remain well defined and predictive when expressed in terms of these sphere-based coordinates.

The basic mechanism that allows for evolution through the big bang in this formalism is clear: when the volume v is recognised as an empirically irrelevant quantity, then the geodesics of the space-time metric lose some of their operational significance. Integrability of the contact equations then replaces geodesic completeness as a criterion for singularity resolution. While some progress has been made in understanding this new criterion more physically (see, for example, Hoffmann & Sloan 2024), ongoing efforts are underway to provide further clarifications. See, however, Koslowski et al. (2018) and Sloan (2019) for the full details of the technical construction and motivations.

13.4 Conclusions

Throughout this work, we have investigated different inequivalent ways in which dynamical systems can be considered closed. These different notions include closure in the sense of conserved energy, conserved measure density, and dynamical autonomy. While we found merits for each, no single notion of closure was completely adequate for all purposes. In particular, we found conventional notions of conserved energy and measure to be wanting when applied to the cosmological setting. In that context, we compared a closed (in the sense of preserved measure density and Hamiltonian) symplectic description of cosmology to an open contact formulation. We found that the latter should be taken to give a better complete description of the system.

Central to our argument was a requirement to treat as physical only those structures that are strictly necessary for maintaining empirical adequacy. The symplectic 'closed' descriptions violate this principle because they treat states as distinct that differ only by the overall scale of the universe (as determined by v), which is not empirically relevant. In contrast, the contact description does satisfy this principle as it excludes v. In other words, while the symplectic and contact systems are both empirically adequate, they are not physically equivalent in the sense that the symplectic system makes definite claims about the value of v—claims which are not necessary for describing cosmological phenomena. While the contact system does not conserve energy or measure density, it is nevertheless autonomous in the empirically sufficient variables, and is thus 'closed' in the third sense defined in the introduction. This suggests that the appropriate notion of closure in the cosmological setting is that of dynamical autonomy in terms of the contact system.

In addition to this epistemic argument favouring the contact description, we also showed that the contact description of cosmology has several other explanatory advantages over the standard symplectic description. First, the contact description provides a natural dynamical explanation for the arrow of time.[5] This arrow comes from the focusing of solutions along the dynamical evolution—particularly when this focusing is monotonic across a dense set of solutions that are consistent with observation. Indeed, we saw that this is precisely what occurs in the cosmological models we considered, since the solution focusing is proportional to the Hubble parameter, and Hubble expansion is monotonic under a dense set of solutions in the presence of a positive cosmological constant.[6] In symplectic description, the natural measure is the time-independent Liouville measure, which exists because the symplectic theory counts states related by different values of v as distinct. It is thus extremely unnatural to use time-dependent measures in the symplectic description. As a result, the arrow of time is usually explained in terms of some version of a past hypothesis. The limitations of a past hypothesis have been discussed by many authors (e.g., Earman 2006, Gryb 2021). In the description we are advocating, these difficulties are avoided because a past hypothesis is unnecessary.

A second advantage of the 'open' model[7] is that, while it agrees with the 'closed' description wherever that description is valid, it continues to be regular when the 'closed' system breaks down. The 'open' model can be integrated through the initial singularity unambiguously. The contact system is thus sufficient to save the phenomena while avoiding the problematic aspects of a big bang singularity.

Endnotes

1. Modulo spatial boundary conditions that could be taken to zero for a closed system.
2. No such finite measure exists in an infinite-dimensional system.
3. We will deal exclusively with classical models in this chapter, sidestepping the important question of how to build a full-fledged quantum model of the early universe. Such models, for instance, may involve modified dynamics on a coarse-grained level.
4. Note that while the numerical value of the Hamiltonian, treated as a phase space function, is preserved, the functional form of its Legendre transform will depend, in general, on the time parameter chosen for the evolution.
5. See Gryb (2025) for more details about which aspects of the problem of the arrow of time can be addressed with this formalism.
6. Particularly when the matter content is required to obey any of the energy conditions of general relativity.
7. Here, we mean 'open' in the sense of having no conserved measure density and Hamiltonian, but 'closed' in the sense of dynamical autonomy.

Bibliography

Ashtekar, A., & Sloan, D. (2010), 'Loop Quantum Cosmology and Slow Roll Inflation', in *Physics Letters B* 694: 108–12.

Ashtekar, A., & Sloan, D. (2011), 'Probability of Inflation in Loop Quantum Cosmology', in *General Relativity and Gravitation* 43: 3619–55.

Belinsky, V. A., Khalatnikov, I. M., & Lifshitz, E. M. (1982), 'A General Solution of the Einstein Equations with a Time Singularity', in *Advances in Physics* 31: 639–67.

Berger, B. K., Garfinkle, D., Isenberg, J., Moncrief, V., & Weaver, M. (1998), 'The Singularity in Generic Gravitational Collapse is Space—Like, Local, and Oscillatory', in *Modern Physics Letters A* 13: 1565–74.

Bravetti, A., Cruz, H., & Tapias, D. (2017), 'Contact Hamiltonian Mechanics', in *Annals of Physics* 376: 17–39.

Bravetti, A., Jackman, C., & Sloan, D. (2023), 'Scaling Symmetries, Contact Reduction and Poincaré's Dream', in *Journal of Physics A: Mathematical and Theoretical* 56: 435203.

Corichi, A., & Karami, A. (2011), 'On the Measure Problem in Slow Roll Inflation and Loop Quantum Cosmology', in *Physical Review D* 83: 104006.

Corichi, A., & Sloan, D. (2014), 'Inflationary Attractors and their Measures', in *Classical and Quantum Gravity* 31: 062001.

Earman, J. (2006), 'The "Past Hypothesis": Not Even False', in *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 37: 399–430.

Foster, S. (1998), 'Scalar Field Cosmologies and the Initial Space-Time Singularity', in *Classical and Quantum Gravity* 15: 3485–504.

Gibbons, G., Hawking, S., & Stewart, J. (1987), 'A Natural Measure on the Set of All Universes', in *Nuclear Physics B* 281: 736–51.

Gryb, S. (2021), 'New Difficulties for the Past Hypothesis', in *Philosophy of Science* 88: 511–32.

Gryb, S. (2025), *Gauge Symmetry and the Arrow of Time: How to Count What Counts*. Phd thesis. University of Groningen.

Gryb, S., & Sloan, D. (2021), 'When Scale is Surplus', in *Synthese* 199: 14769–820.

Heinzle, J. M., & Uggla, C. (2009), 'Mixmaster: Fact and Belief', in *Classical and Quantum Gravity* 26: 075016.

Hoffmann, J., & Sloan, D. (2024), 'Continuation of Bianchi Spacetimes Through the Big Bang', in *Physical Review D* 110: 063539.

Koslowski, T. A., Mercati, F., & Sloan, D. (2018), 'Through the Big Bang: Continuing Einstein's Equations Beyond a Cosmological Singularity', in *Physics Letters B* 778: 339–43.

Lazo, M. J., Paiva, J., Amaral, J. T. S., & Frederico, G. S. F. (2018), 'An Action Principle for Action-Dependent Lagrangians: Toward an Action Principle to Non-Conservative Systems', in *Journal of Mathematical Physics* 59: 032902.

Mercati, F. (2019), 'Through the Big Bang in Inflationary Cosmology', in *Journal of Cosmology and Astroparticle Physics* 10: 025.

Ringstrom, H. (2001), 'The Bianchi IX Attractor', in *Annales Henri Poincaré* 2: 405–500.

Sloan, D. (2018), 'Dynamical Similarity', in *Physical Review D* 97: 123541.

Sloan, D. (2019), 'Scalar Fields and the FLRW Singularity', in *Classical and Quantum Gravity* 36: 235004.

Sloan, D. (2021), 'New Action for Cosmology', in *Physical Review D* 103: 043524.

Sloan, D. (2023), 'Herglotz Action for Homogeneous Cosmologies', in *Classical and Quantum Gravity* 40: 115008.

Sloan, D., & Silk, J. (2016), 'Volume Weighting the Measure of the Universe from Classical Slow-Roll Expansion', in *Physical Review D* 93: 104030.

Uggla, C. (2007). 'The Nature of Generic Cosmological Singularities', in *Recent Developments in Theoretical and Experimental General Relativity, Gravitation and Relativistic Field Theories. Proceedings, 11th Marcel Grossmann Meeting, MG11, Berlin, Germany, July 23-29, 2006. Pt. A-C* (World Scientific), 73–89.

14
Density Matrix Realism

Eddy Keming Chen

14.1 Introduction

Realism about quantum theory naturally leads to realism about the quantum state of the universe, the idea that the universal quantum state is an objective and mind-independent feature of the physical world. It leaves open, however, whether it is a pure state represented by a wave function, or an impure (mixed) one represented by a density matrix. In this chapter, I characterise and elaborate on Density Matrix Realism (DMR), the thesis that the universal quantum state is objective but *can be impure*.

Related ideas have been considered in quantum foundations (Dürr et al. 2005, Maroney 2005, Tumulka 2010, Allori et al. 2013, Wallace 2012, Chen and Tumulka 2022), quantum cosmology (Page 1986, Page 2008; Barvinsky and Kamenshchik 2006), quantum gravity (Hawking 1976, 1982), and philosophy of science (McCoy 2020, Robertson 2022, Cuffaro and Hartmann 2021).

DMR was formulated and given its name in Chen (2018),[1] with further developments in Chen (2020; 2022a; 2022b; 2023a; 2024b; 2024c). DMR provides a broad (and I think fruitful) framework for thinking about quantum theory.

For various reasons, DMR remains puzzling to many people. The common issues are:

1. Density matrices are often regarded as second-class citizens in quantum theory, representing subsystem states (in the form of reduced density matrices) or epistemic states (in the form of statistical density matrices). Such a perspective is taught in standard textbooks on quantum mechanics and assumed in many research articles in quantum foundations.
2. Density-matrix theories of quantum mechanics are less familiar than wave-function theories. Although both can be adapted to solve the quantum measurement problem, the density-matrix solutions are not widely known.
3. It is unclear how to interpret the density matrix as an objective feature of the universe.
4. It is unclear whether and how density-matrix theories can reproduce the same empirical predictions as wave-function theories. In other words, it is

Eddy Keming Chen, *Density Matrix Realism*. In: *Open Systems: Physics, Metaphysics, and Methodology*.
Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press. © Oxford University Press (2026).
DOI: 10.1093/9780198929277.003.0015

unclear whether and under what conditions such theories are empirically equivalent.
5. It is unclear what theories with a universal density matrix look like from the subsystem level.
6. There are multiple versions of DMR; it is sometimes unclear which one is being discussed in a particular context.

This chapter attempts to address such issues. First, I discuss some definitions, examples, and interpretations of density-matrix realist theories. Second, I suggest that wave-function theories and density-matrix theories can be empirically equivalent at the universal level, the subsystem level, and the local level. Third, I consider two generalisations of Density Matrix Realism. Finally, I answer six frequently asked questions and highlight an implication for scientific realism.

The goal here is to clarify DMR. There are other issues worth considering, regarding the motivations for and arguments in favour of DMR. Although they are not the focus here, I briefly address them in §14.7.6. See the companion paper Chen (2024c) for more related discussions and references.

14.2 Definitions

I begin by defining Wave Function Realism (WFR), a more familiar and popular thesis about quantum state realism, and a close cousin to DMR.

Consider realist solutions to the quantum measurement problem, such as Bohmian mechanics, objective collapse theories, and Everettian quantum mechanics. In such theories, the universal quantum state plays an indispensable role in the kinematics and the dynamics, providing reasons to regard it as an objective feature of reality. Given how such theories are often presented, it is widely believed that the objective universal quantum state has to be pure, represented by a wave function. Let us formulate the thesis as follows:

Wave Function Realism (WFR) The quantum state of the universe is objective; it has to be pure.

This characterisation of WFR is broader than that of Albert (1996) and Ney (2021). For them, WFR carries a specific commitment to understanding the universal wave function as a physical field that lives on a vastly high-dimensional 'configuration' space, from which the ordinary 3-dimensional space is emergent. That is, however, not the only way to be a realist about the wave function. For example, the multi-field interpretation, spacetime state realism, and the nomological interpretation all count as realist interpretations of the wave function, and hence properly fall under the same label (Chen 2019b).

WFR, thus defined, is a standard position for realism about the universal quantum state. We often assume that the universal quantum state, if objective, must be pure. In textbook presentations, mixed states are often used to represent reduced or statistical states, as expressions of entanglement with other systems or our ignorance of the actual pure state. However, there is no compelling reason why the universe cannot be in a fundamental mixed state, one that does not arise from entanglement or lack of knowledge. In fact, we can formulate Bohmian mechanics, objective collapse theories, and Everettian quantum mechanics with a fundamental density matrix (Dürr et al. 2005, Maroney 2005, Allori et al. 2013, Wallace 2012). I explain them in §14.3.

Let us consider an alternative to WFR:

Density Matrix Realism (DMR) The quantum state of the universe is objective; it can be pure or impure.

This thesis provides additional theoretical freedom in choosing the quantum state of the universe; it can be pure or impure (a 'mixed state'). The option to use impure states is crucial for certain proposals in quantum foundations and quantum cosmology (§14.3).

Let us clarify some key terms in both theses.

(i) 'The': it implies uniqueness. Both theses differ from the proposal that seems to be suggested by Wallace (2016) that the universe can possess two physical states at the same time: a fundamental pure state and a fundamental mixed state.

(ii) 'Quantum state of the universe': both theses are about the quantum state of the universe. This does not logically entail that subsystem quantum states must be objective or that they must be pure (or impure). We consider a generalisation to subsystem quantum states in §14.6.

(iii) 'Objective': it means that the universal quantum state corresponds to an objective feature of reality, which is not merely epistemic (encoding lack of knowledge), or pragmatically useful (merely an instrument for calculations). The meaning of objectivity is left open-ended, leaving room for different realist interpretations (Chen 2019b). We consider several options in §14.4.

(iv) 'Must be pure' vs. 'can be pure or impure': this is the only difference between the two theses. 'Must' and 'can' are modal concepts. WFR restricts universal quantum states to only pure ones, while DMR allows both pure and impure universal quantum states. However, the latter is compatible with additional laws of physics (such as the Initial Projection Hypothesis proposed in Chen 2018) that make it physically impossible for the universe to be in a pure state.

14.3 Examples

Let us illustrate WFR and DMR with some examples of physical theories.

Ψ-BM. Consider the standard formulation of Bohmian mechanics (Ψ-BM) that validates WFR. The state of the universe at a time is given by (Ψ, Q). The universal wave function Ψ that evolves unitarily according to the Schrödinger equation:

$$i\hbar\frac{\partial\Psi}{\partial t} = H\Psi. \tag{1}$$

There are actual particles that have precise locations in physical space, represented by $\mathbb{R}^3$. The change in particle configuration $Q = (Q_1, Q_2, ..., Q_N) \in \mathbb{R}^{3N}$ follows the guidance equation (written for the i-th particle):

$$\frac{dQ_i}{dt} = \frac{\hbar}{m_i}\mathrm{Im}\frac{\nabla_i\Psi(q)}{\Psi(q)}(q = Q). \tag{2}$$

Moreover, the initial particle distribution is given by the quantum equilibrium distribution:

$$\rho_{t_0}(q) = |\Psi(q, t_0)|^2. \tag{3}$$

W-BM. Consider an alternative formulation of Bohmian mechanics (W-BM) whose natural interpretation validates DMR. The state of the universe at a time is given by (W, Q). The universal density matrix W (which can be a mixed state) evolves unitarily according to the von Neumann equation:

$$i\hbar\frac{\partial W}{\partial t} = [H, W]. \tag{4}$$

The particle configuration $Q = (Q_1, Q_2, ..., Q_N) \in \mathbb{R}^{3N}$ follows a new guidance equation (Dürr et al. 2005):

$$\frac{dQ_i}{dt} = \frac{\hbar}{m_i}\mathrm{Im}\frac{\nabla_{q_i}W(q, q', t)}{W(q, q', t)}(q = q' = Q) \tag{5}$$

and is initially distributed by

$$\rho_{t_0}(q) = W(q, q, t_0)dq. \tag{6}$$

This version of Bohmian mechanics satisfies equivariance just as the wave-function version does (Dürr et al. 1992, Dürr et al. 2005).

Ψ-GRW. There are several versions of GRW theories that validate WFR. In the first one, Ψ-GRW0, the fundamental ontology consists only of the universal wave

function. The wave function typically obeys (1), but the linear evolution is interrupted randomly (with rate $N\lambda$, where N is the number of particles and λ is a new constant of nature of order 10^{-15} s^{-1}) by collapses:

$$\Psi_{T^+} = \frac{\Lambda_k(X)^{1/2}\Psi_{T^-}}{||\Lambda_k(X)^{1/2}\Psi_{T^-}||} \tag{7}$$

where Ψ_{T^-} is the pre-collapse wave function, Ψ_{T^+} is the post-collapse wave function, the collapse centre X is chosen randomly with probability distribution $\rho(x) = ||\Lambda_k(x)^{1/2}\Psi_{T^-}||^2 dx$, $k \in \{1, 2, \ldots, N\}$ is chosen randomly with uniform distribution on that set of particle labels, and the collapse rate operator is defined as:

$$\Lambda_k(x) = \frac{1}{(2\pi\sigma^2)^{3/2}} e^{-\frac{(Q_k - x)^2}{2\sigma^2}} \tag{8}$$

where Q_k is the position operator of 'particle' k, and σ is another new constant of nature of order 10^{-7} m postulated in current GRW theories. We can also define Ψ-GRWf and Ψ-GRWm, a flashy version and a mass-density version of GRW with distinct choices of the local beables. For example, Ψ-GRWm includes an additional law for the mass-density ontology in spacetime:

$$m(x, t) = \langle \Psi(t) | M(x) | \Psi(t) \rangle \tag{9}$$

where x is a physical space variable, $M(x) = \sum_i m_i \delta(Q_i - x)$ is the mass-density operator, which is defined via the position operator $Q_i \psi(q_1, q_2, \ldots, q_n) = q_i \psi(q_1, q_2, \ldots, q_n)$.

W-GRW. Consider a class of alternative formulations of GRW that validate DMR. For W-GRW0, the fundamental ontology consists only of the universal density matrix W, which typically obeys (4), but the linear evolution is interrupted randomly:

$$W_{T^+} = \frac{\Lambda_{I_k}(X)^{1/2} W_{T^-} \Lambda_{I_k}(X)^{1/2}}{\mathrm{tr}(W_{T^-} \Lambda_{I_k}(X))} \tag{10}$$

with W_{T^+} the post-collapse density matrix, W_{T^-} the pre-collapse density matrix, and X distributed by the probability density $\rho(x) = \mathrm{tr}(W_{T^-} \Lambda_{I_k}(x))$.

We can similarly define W-GRWf and W-GRWm, the flashy version and the mass-density version (Allori et al. 2013). For example, W-GRWm includes the mass-density law:

$$m(x, t) = \mathrm{tr}(M(x) W(t)). \tag{11}$$

Ψ-EQM. In the standard version of Everettian quantum mechanics with a fundamental pure state, the universal wave function evolves unitarily by (1), with the

branching structure emergent from decoherence. We also have the option to add a separable fundamental ontology in spacetime by defining a mass-density field as in (9).

W-EQM. In the alternative formulation of Everettian quantum mechanics with a fundamental density matrix, we can unitarily evolve the universal density matrix W by (4), understand the emergent branching structure via decoherence, and apply self-locating probability to recover the Born rule (Chua and Chen 2025). We also have the option to add a separable fundamental ontology in spacetime by defining a mass-density field as in (11).

Such theories illustrate the similarities and differences between WFR and DMR. The Ψ-theories and the W-theories employ similar equations, but the fundamental objects are different. The Ψ-theories postulate a universal wave function, while the W-theories postulate a universal density matrix that may be a fundamental mixed state.

For both WFR and DMR, we can postulate other laws to further restrict the nomologically possible quantum states. For example, to explain the manifest arrows of time in Ψ theories, we may postulate a Past Hypothesis (PH) to constrain the initial wave functions to those compatible with $\mathscr{H}_{PH}$, a low-dimensional subspace corresponding to the initial low-entropy macrostate of the universe. W-theories offer some distinctive choices. A particularly natural one is to apply the Initial Projection Hypothesis (IPH) and pick the normalised projection onto $\mathscr{H}_{PH}$ as the only nomologically possible initial quantum state (Chen 2018, Chen 2024c). Such a choice requires DMR (as long as $\mathscr{H}_{PH}$ has more than one dimension).

As another example, consider the on-going discussion in quantum cosmology inspired by the No-Boundary Proposal (Hartle and Hawking 1983). In the WFR framework, we can follow the Hartle and Hawking prescription (insofar as it is well-defined) and select the universal wave function to be the one where space smoothly shrinks to a point towards one temporal boundary. However, there are also proposals that require DMR, such as Page's No-Bang Cosmology (Page 2008) and the thermal version of the Hartle–Hawking state (Barvinsky and Kamenshchik 2006). Certain approaches to the black hole information loss paradox require DMR (Hawking 1976, Hawking 1982), but they remain controversial. DMR may also be a natural setting (but not a requirement) for investigating the thermal time hypothesis (Connes and Rovelli 1994), according to which the time variable is determined by a universal density matrix.

14.4 Interpretations

What does a universal density matrix represent in the physical world? The density matrix is an abstract mathematical object. It can be challenging to say what kind of thing it represents. Nevertheless, the situation is no more difficult than that in

the case of WFR. Strategies for interpreting the universal wave function are also available for interpreting the universal density matrix.

Similar to the situation in WFR (Chen 2019b), we have four ontological interpretations of the universal density matrix in DMR. First, we can understand $W(q, q', t)$ as representing a physical field evolving in a $6N$-dimensional fundamental space represented by $\mathbb{R}^{6N}$. The field assigns properties to every point on that space. Second, we can understand it as representing a low-dimensional multi-field. The fundamental space is a three-dimensional space represented by $\mathbb{R}^3$, and $W(q, q', t)$ assigns properties to every $2N$-tuple of points on that space. Third, we can understand it as describing properties of spacetime regions. We can obtain, from the universal density matrix, reduced density matrices that correspond to physical properties of regions in a four-dimensional manifold. Such properties are in general non-separable due to quantum entanglement. Finally, we can understand it as representing a geometric object in Hilbert space.

For the ontological interpretations, locality and separability considerations always go together in WFR but may come apart in DMR. For example, in W-BM, the smallest space where everything is separable is $\mathbb{R}^{6N}$, while the smallest space where the dynamics is local is $\mathbb{R}^{3N}$, since the particle configuration lies in $\mathbb{R}^{3N}$. Hence, for those who understand the quantum state as a field-like material object (Albert 1996, Ney 2021), they face a choice on whether the fundamental physical space is $3N$-dimensional or $6N$-dimensional.

Alternatively, we can adopt the nomological interpretation of the quantum state. In W-BM, we can regard W_0 as a function that prescribes the velocity field for the particle configuration, just like the Hamiltonian function prescribes the velocity field on phase space in classical mechanics. In W-EQM with a mass-density ontology, we have the option of postulating the $m(x, t)$ field as the only material ontology, whose time evolution is given by a law written in terms of a mathematical function corresponding to a universal density matrix. DMR allows us to choose a simple initial density matrix (Chen 2018, Chen 2022a). So we have an easy route to the nomological interpretation of the universal quantum state in DMR.

14.5 Empirical Equivalence

In §14.3 we saw some examples of physical theories with fundamental universal density matrices that are natural to understand in the DMR framework. Given appropriate choices of physical laws, each density-matrix theory is empirically equivalent to its wave-function counterpart, so that they cannot be distinguished even in principle by experiment or observation (Chen 2019a).

This is important, for it means the two kinds of theories rise and fall together with respect to empirical evidence. If one is confirmed by evidence, the other is too, and by the same degree. Conversely, if one is disconfirmed by evidence, so

is the other. However, the relevant statement of empirical equivalence depends on not just the quantum state but also the laws governing its history, and this is a place where different choices for the density-matrix dynamics can make a difference. In §14.6, we will consider an empirical difference in the generalised quantum theory advocated by Cuffaro and Hartmann (2021).

There are different ways to understand the empirical equivalence of theories. I shall approach it from three levels. The first is at the universal level, the second at the subsystem level, and the third at the local level, with significant overlap between the latter two. Even though wave-function realist theories and density-matrix realist theories are physically distinct, they can be empirically equivalent at all three levels.

14.5.1 The Universal Level: Equivalence of Probabilities

We start with a general conception of empirical equivalence:

Criterion for Empirical Equivalence: Theories A and B are empirically equivalent if they assign the same probability to every possible outcome in every possible measurement.

The criterion should be understood at the universal level, since it concerns all possible measurements that can be done in the universe. If two theories satisfy this criterion, they are empirically equivalent in a very strong sense, concerning not just actual data, but all possible data; not just the actual world, but all possible worlds, permitted by their physical laws. No observation or experiment can distinguish between them.

The general criterion can take on different forms in different interpretations of quantum mechanics. For example, we can formulate a special case for Bohmian theories. Since in a Bohmian universe every measurement apparatus is made out of particles with precise positions, every measurement boils down to a position measurement. Hence, to achieve the strong sense of empirical equivalence, we just need two theories to agree on the probability of the particle configurations.

Criterion for Empirical Equivalence of Bohmian Theories: Bohmian theories A and B are empirically equivalent if, for every time t, $P_A(Q_t \in dq) = P_B(Q_t \in dq)$.

We can formulate W-BM and Ψ-BM in such a way that the criterion is satisfied.

Theorem 14.5.1 *Let W-BM be the theory of* (W, Q) *such that* W *evolves by (4) and* Q *evolves by (5) and satisfies (6); moreover, a particular* $W(t_0)$ *is chosen. Let* Ψ*-BM be the theory of* (Ψ, Q) *such that* Ψ *evolves by the Schrödinger*

equation, Q evolves by the guidance equation and satisfies the quantum equilibrium distribution in the wave-function version of Bohmian mechanics; moreover, $\Psi(t_0)$ is chosen at random from a statistical ensemble represented by the density matrix $W(t_0)$. W-BM and Ψ-BM are empirically equivalent in the sense above.

Proof: see Chen (2019a). The basic idea is that, because of equivariance, at every time, the two theories have the same formula for the probability distribution of the particle configuration.

For Everettian theories, we can draw the same conclusion about the respective theories. Let W-EQM be the theory of W such that W evolves by (4); moreover, a particular $W(t_0)$ is chosen. Let Ψ-EQM be the theory of Ψ such that Ψ evolves by (1); moreover, $\Psi(t_0)$ is chosen at random from a statistical ensemble represented by the density matrix $W(t_0)$.

W-EQM and Ψ-EQM are empirically equivalent, because they assign the same probability to every outcome in every experiment. Let A be a self-adjoint operator corresponding to some observable. Suppose that its spectral measure is given by $\mathscr{A}$, a projection-valued measure. Then the probability that at time t, the outcome of the measurement x will be within some measurable set M is:

- W-EQM: $P_{W\text{-}EQM}(x \in M) = \mathrm{tr}(W_t \mathscr{A}(M))$.
- Ψ-EQM: $P_{\Psi\text{-}EQM}(x \in M) = \mathrm{tr}(W_t \mathscr{A}(M))$.

Under the assumption that probability makes sense in the Ψ-Everettian and W-Everettian theories, the two types of Everettian theories are empirically equivalent.

For GRW theories, the W and Ψ versions with different choices of the primitive ontology (m or f) are also empirically equivalent to each other (with the arrows denoting empirical equivalence):

$$\begin{array}{ccc} \Psi\text{-GRWf} & \longleftrightarrow & W\text{-GRWf} \\ \updownarrow & & \updownarrow \\ \Psi\text{-GRWm} & \longleftrightarrow & W\text{-GRWm} \end{array} \tag{12}$$

For arguments for why they are empirical equivalent, see Allori et al. (2013) and Chen (2019a).

14.5.2 The Subsystem Level: Collapses

The general argument for empirical equivalence is clean and powerful. Nevertheless, it may be useful to analyse how the universe works at the subsystem level and how WFR and DMR can converge on their predictions for the subsystem dynamics. In this section, I shall use Bohmian mechanics as a concrete example, for its

sharp mathematical structure offers a particularly clear picture of how subsystem states can differ on Ψ-BM and W-BM but lead to the same empirical consequences.

Following Chen (2019a), consider the following analysis of subsystems in *W*-BM:

(1) **Splitting.** For any given subsystem of particles we have a splitting:

$$q = (x, y), \tag{13}$$

with x the generic variable for the configuration of the subsystem and y the generic variable for the configuration of the environment, i.e., the complement of the subsystem. (13) provides a splitting of the actual configuration into two parts:

$$Q = (X, Y). \tag{14}$$

So we can write the universal density matrix in terms of $W = W(x, y, x', y')$.

(2) **Effective density matrix.** The subsystem corresponding to the x-variables has an *effective density matrix* (at a given time) if the universal density matrix $W(x, y, x', y')$ and the actual configuration $Q = (X, Y)$ (at that time) satisfy:

$$W(x, y, x', y') = \rho(x, x')\gamma(y, y') + W^{\perp}(x, y, x', y'), \tag{15}$$

such that $\gamma(y, y')$ and $W^{\perp}(x, y, x', y')$ have macroscopically disjoint (y, y)-supports, and

$$(Y, Y) \in \text{supp } \gamma(y, y'), \tag{16}$$

In this case, the effective density matrix of the subsystem is $\rho(x, x')$.

(3) **Conditional density matrix.** The effective density matrix for a subsystem does not always exist. However, we can always define the *conditional density matrix* in the following way:

$$w(x, x') = W(x, Y, x', Y). \tag{17}$$

Here we identify quantum states differing by a constant factor. Given the definition of the velocity (5), the velocity field of the x-system will be given by its conditional density matrix.

(4) **Collapse and effective collapse.** When (15) and (16) are satisfied, we can neglect, for all practical purposes, $W^{\perp}(x, y, x', y')$. The configuration will be carried by the relevant part of the universal density matrix—$\rho(x, x')\gamma(y, y')$—into the future, without much interference from the other parts contained in $W^{\perp}(x, y, x', y')$. In this case, we can say that during measurement, the universal density matrix has undergone an effective collapse from W_{t^-} to

$W_{t^+} = \rho(x, x', t^+)\gamma(y, y', t^+)$. Nevertheless, for the subsystem density matrix, represented by the conditional density matrix $w(x, x', t)$, the time evolution is not given by a von Neumann equation but will in measurement-like situations look practically like a collapse, with a rapid evolution to $w(x, x', t^+)$.

(5) **The Fundamental Conditional Probability Formula.** By equivariance, the distribution of Q_t is always given by $W(q, q, t)$. By (17), at time t, for the conditional probability distribution of the configuration of a subsystem X_t given the actual configuration of the environment Y_t, we have the fundamental conditional probability formula for W theories:

$$P(X_t \in dx|Y_t) = w(x, x, t)dx, \tag{18}$$

where $w(x, x', t) = w(x, x', t)^{Y_t}$ is the conditional density matrix of the subsystem at time t. Similar to the situation in Ψ-BM, the configurations X_t and Y_t are conditionally independent given the density matrix $w(x, x', t)$.

It is easy to check that pure effective density matrices can emerge following collapses on the conditional states (see Chen 2019a for a simple example). What is surprising is that W-BM offers a new type of collapses that have no analogue in textbook quantum mechanics or even Ψ-BM. The following example shows in a concrete way how subsystem dynamics can differ on the two theories and yet produce the same probabilistic predictions.[2]

Consider two momentum eigenstates, one with momentum +1 (moving uniformly to the right) and the other −1 (moving uniformly to the left). On Ψ-BM, suppose one of them is the actual quantum state guiding the particle. The particle will be either moving to the left or to the right. However, on W-BM, if we regard the equal mixture of the two as the fundamental density matrix, one might worry that the particle guided by this density matrix will be permanently at rest, because the average of the two gives exactly zero momentum. Nevertheless, when we measure the particle, the record may indicate that the particle is moving. What is going on?

Let us analyse the measurement situation with the subsystem analysis provided above. Let us split the universal configuration Q into (X, Y), with X the subsystem configuration and Y the environmental one. At t_1, before measurement, consider two possible universal 'wave functions' (setting aside the issue of square-integrability):

$$\Psi_A^{t_1}(x, y) = \psi^-(x)\phi^{ready}(y) \tag{19}$$

$$\Psi_B^{t_1}(x, y) = \psi^+(x)\phi^{ready}(y) \tag{20}$$

where $\psi^-(x)$ and $\psi^+(x)$ denote the momentum eigenstates with momentum −1 and momentum +1 and $\phi^{ready}(y)$ the quantum state of the environment that is ready

to measure the particle's momentum. Let us suppose the two will unitarily evolve into these:

$$\Psi_A^{t_2}(x,y) = \psi^-(x)\phi^-(y) \tag{21}$$

$$\Psi_B^{t_2}(x,y) = \psi^+(x)\phi^+(y) \tag{22}$$

with $\phi^-(y)$ and $\phi^+(y)$ the states of the environment indicating the particle in the subsystem is moving to the left and moving to the right, respectively. As usual, we assume that the two records are macroscopically distinct, with $\phi^-(y)$ and $\phi^+(y)$ having mathematically disjoint supports.

Consider an impure density matrix, the equal mixture of $\Psi_A^{t_1}(x,y)$ and $\Psi_B^{t_1}(x,y)$:

$$W^{t_1}(x,y,x',y') = \frac{1}{2}\Psi_A^{t_1}(x,y)\Psi_A^{t_1*}(x',y') + \frac{1}{2}\Psi_B^{t_1}(x,y)\Psi_B^{t_1*}(x',y') \tag{23}$$

Let the universal particle configuration be guided by such a density matrix according to the W-BM dynamics.

Before measurement, at t_1, the conditional density matrix of the one-particle subsystem is an impure state:

$$w^{t_1}(x,x') = KW^{t_1}(x,Y^{t_1};x',Y^{t_1}) = \frac{1}{2}\psi^-(x)\psi^{-*}(x') + \frac{1}{2}\psi^+(x)\psi^{+*}(x') \tag{24}$$

with K the normalisation constant.

After measurement, at t_2, the subsystem is measured and the record indicates that it is actually moving to the left. The universal quantum state has evolved into this:

$$W^{t_2}(x,y,x',y') = \frac{1}{2}\Psi_A^{t_2}(x,y)\Psi_A^{t_2*}(x',y') + \frac{1}{2}\Psi_B^{t_2}(x,y)\Psi_B^{t_2*}(x',y'). \tag{25}$$

Because of the macroscopically disjoint supports of $\phi^-(y)$ and $\phi^+(y)$, the two terms on the right-hand side, $\Psi_A^{t_2}(x,y)\Psi_A^{t_2*}(x',y')$ and $\Psi_B^{t_2}(x,y)\Psi_B^{t_2*}(x',y')$, also have macroscopically disjoint supports.

At t_2, the environmental configuration Y^{t_2} lies within the support of $\phi^-(y)$. Plugging it into the universal density matrix, we have (almost) zero contributions from terms with $\phi^+(y)$. So the conditional density matrix of the subsystem becomes a pure state:

$$w^{t_2}(x,x') = K'W^{t_2}(x,Y^{t_2};x',Y^{t_2}) = \psi^-(x)\psi^{-*}(x') \tag{26}$$

with K' the normalisation constant. The evolution of the conditional quantum state of the subsystem from $w^{t_1}(x,x')$ into $w^{t_2}(x,x')$ is a non-unitary one, corresponding to a more general type of quantum state collapses. The collapse into

$\psi^{-}(x)\psi^{-*}(x')$ has a probability of 1/2, the same as the collapse into $\psi^{+}(x)\psi^{+*}(x')$. Notice that the collapse *takes a mixed state to a pure state*. It is a real collapse for the subsystem state in W-BM. However, this is a case where textbook QM and Ψ-BM do not postulate a collapse at all; they would regard the transition as an epistemic change of learning which of the two pure states is the actual one, or, in Bayesian terms, update by conditionalisation. Let us call this '*collapse without textbook collapse*'. It illustrates a crucial difference between Ψ-BM and W-BM at the subsystem level. In this case, the conditional state collapses in W-BM but the corresponding one does not collapse in Ψ-BM, and yet they are predictively equivalent, assigning the same probabilities to possible outcomes. In other words, they offer different physical explanations for the empirical phenomena.

14.5.3 The Local Level: Canonical Typicality

The previous arguments are quite general and sufficient to establish exact empirical equivalence with exact agreements of probabilities. I shall propose another argument for an approximate sort of empirical equivalence between WFR and DMR based on local indistinguishability of 'small' subsystems in a 'large' universe. It exploits recent results in quantum statistical mechanics known as *canonical typicality*. Its scope is more restricted: the argument depends on the size of the actual universe and how much of it we have direct empirical access to. However, because of the generous bound on subsystem size, it presumably applies to our actual epistemic situation. I include the argument here because it is surprising and interesting.

In the early 2000s, several teams of researchers (Gemmer and Mahler 2003, Goldstein et al. 2006, Popescu et al. 2006) independently discovered a feature of quantum mechanics that has no parallel in classical mechanics. Suppose we have a large quantum universe U with a wave function Ψ, with the universe partitioned into a small subsystem S and its environment E (the complement of S in U) and the universal Hilbert space is the tensor product of the Hilbert spaces of S and E, i.e., $\mathscr{H} = \mathscr{H}_S \otimes \mathscr{H}_E$. Suppose further that there is a global constraint R on the universe such that $\Psi \in \mathscr{H}_R$, with $\mathscr{H}_R$ the subspace of the universal Hilbert space corresponding to R. The reduced density matrix of the subsystem S, obtained from Ψ by tracing out the environmental degrees of freedom in E, is

$$\rho_S^{\Psi} = \mathrm{tr}_E|\Psi\rangle\langle\Psi| \tag{27}$$

Now, the normalised projection onto $\mathscr{H}_R$ is $W_R = \mathbb{I}_R/\dim \mathscr{H}_R$, with $\mathbb{I}_R$ the projection onto $\mathscr{H}_R$. W_R is mathematically the same as the statistical density matrix corresponding to the uniform probability distribution, with respect to the normalised

surface area measure μ, on the unit sphere of $\mathscr{H}_R$:

$$W_R = \int_{\mathbb{S}(\mathscr{H})} \mu(d\psi)\, |\psi\rangle\langle\psi| \tag{28}$$

The reduced density matrix of the subsystem S, obtained from W_R by tracing out the environmental degrees of freedom in E, is

$$\rho_S^{W_R} = \mathrm{tr}_E W_R \tag{29}$$

Remarkably, it is shown that, under suitable conditions and for typical wave functions in $\mathscr{H}_R$,

$$\rho_S^{\Psi} \approx \rho_S^{W_R} \tag{30}$$

where the approximation can be made precise by various inequalities (Teufel et al., 2025). This is known as canonical typicality. The phenomenon is a manifestation of quantum entanglement and has no classical analogue.

Popescu et al. (2006) gloss canonical typicality as almost every pure state of the universe is locally (that is, on the system) indistinguishable from W_R. They interpret the latter as encoding a statistical density matrix. Results of this form have been used to support the individualistic picture of quantum statistical mechanics that it is meaningful to use standard statistical ensembles for subsystem descriptions even when the universe is in a pure state, and to justify the postulate of micro-canonical ensembles and canonical ensembles.

However, there is an under-appreciated corollary of canonical typicality, relevant to the empirical equivalence between WFR and DMR. Precisely because of canonical typicality, we cannot use local data to distinguish whether the universal quantum state is pure or impure.

How close ρ_S^{Ψ} and $\rho_S^{W_R}$ are to each other depends on the size of the subsystem and the size of the environment. To apply canonical typicality, the subsystem can contain up to 50% of the number of degrees of freedom of the total system (e.g., the universe) (Teufel et al. 2025). Let's consider the solar system as such a 'small' subsystem of the universe. The spatial region of the solar system includes all our direct evidence of the universe collected so far. We may have other observational data about the universe, such as other galaxies, but it will be inferred from our direct evidence (pointer readings), which is locally here in the solar system. Given the local data in the solar system, we only have access to the reduced density matrix of the solar system. The latter, by canonical typicality, is more or less the same whether it is obtained from a typical wave function in the present macrostate of the universe $\mathscr{H}_M$ or the normalised projection onto the subspace associated with the macrostate, a mixed state. Therefore, given the local data we have, we cannot tell whether the universe is in a pure state or a mixed state.[3]

We are gradually exploring other parts of the universe. But it is realistic to expect that we will not have direct empirical access to more than 50% of the microscopic degrees of freedom in the universe. Considered as an abstract subsystem, our evidence by itself will (typically) not be able to distinguish between a universal pure state or a universal mixed state. In other words, the totality of direct empirical evidence, as long as it is sufficiently local, will typically (according to the standard measure) underdetermine the universal quantum state. This argument requires that our evidence be local in the appropriate sense. As it turns out, by appealing to another result called *distribution typicality*, we can drop the locality assumption and generalise the argument to any empirical evidence (Chen and Tumulka 2024).

14.6 Generalisations

In §14.3 and §14.5, I have considered density-matrix realist theories that are 'conservative extensions' from their wave-function realist counterparts. The laws in the former are a natural generalisation of those in the latter. For example, (4) conservatively extends from (1), (5) from (2), and (10) from (7). All of those laws may be called closed-system dynamics. They differ from the more general open-system dynamics suggested by Cuffaro and Hartmann (2021). Regarding my definition of DMR in §14.2, the relevant quantum state whose objectivity is at issue is the universal one. With the focus on closed-system dynamics and universal quantum states, we can establish the empirical equivalence between WFR and DMR. For clarity, let me label the version of DMR I am committed to as DMR^U_C, with U standing for the universe and C closed-system dynamics.

In this section, let us consider two generalisations of DMR. They are related to the proposals of Robertson (2022) and Cuffaro and Hartmann (2021). There are important similarities and differences.

14.6.1 The Subsystem Version: DMR^S

First, consider DMR^S, the thesis that the reduced density matrices assigned to subsystems of the universe are objective. This view faces an apparent objection. Since reduced states are obtained by tracing out the environment, which is akin to taking a statistical average, they surely cannot be entirely objective. The standard view regards the reduced states as carrying incomplete information about subsystems. Against the standard view, Robertson (2022) argues for DMR^S by suggesting that density matrices are inevitable and the most complete description for subsystems entangled with their environments, concluding that density matrices should not be understood as entities that encode subjective information but as states of the relevant physical systems:[4]

> The density matrix is arguably the best mathematical object to represent the state of the individual system (rather than the wavefunction ψ), since ρ is a more general object than ψ. Quantum systems rapidly become entangled with their environment—which means that the individual system cannot be described by a wavefunction, but instead must be a (reduced) density matrix (by tracing over the environment). Since the density matrix formalism is more general, and sometimes required, the density matrix should be taken to be more fundamental ... Thus the individual state of the system in QM is not represented by a ray in Hilbert space (the quantum equivalent of a point in phase space), but a density matrix ... Thus, there is no difference in the mathematical object that represents the state of the individual system, and a probability distribution over it. Thus, in QSM, the dichotomy between 'being a property of a probability distribution' and 'being a property of the individual system' never arises (ibid., 1013).

But the defender of the standard view has a ready response: the dichotomy can still arise in quantum statistical mechanics. For example, even though the hydrogen atom is entangled with the rest of the universe, we attribute to it a pure state, while the reduced density matrix is heavily mixed. How different quantum theories explain this phenomenon is briefly discussed in §14.5.3. If WFR is correct, in Bohmian mechanics the subsystem can possess a (conditional) pure state, obtained from plugging the environmental configuration into the universal wave function. Even on DMR^U_C, such a conditional state obtained from the universal density matrix can still be pure. (The Everettian story will involve branching.)

Hence, there is still a choice. Empirically equivalent but physically distinct theories can assign different subsystem states, some pure and some mixed. One has the option to regard the mixedness as a symptom for non-objectivity. Hence, we are not forced to accept DMR^S.

However, if we accept DMR^U_C, there is a natural argument for a restricted version of DMR^S. If the universal quantum state is mixed, the actual states of *some* subsystems will be more mixed than when we assume WFR. Equation (24) in §14.5.2 provides an example. See Dürr et al. (2005, sect. 4) for another type of example involving spin. Such mixed states can be objective, though not as fundamental as the universal density matrix. Hence, we may have motivations to adopt the restricted version of DMR^S, treating some subsystem density matrices as objective (though not fundamental), if we already accept DMR^U_C.

14.6.2 The Open Systems View: DMR^U_O

Next, consider DMR^U_O, the idea that the universal density matrix is objective, can be impure, and evolves non-unitarily according to a fundamental open-system dynamical equation, such as the Lindblad equation (Cuffaro and Hartmann 2021).

For example, one can formulate a Bohm-type theory where particles exist in three-dimensional space and are guided by a universal density matrix that evolves according to, not the von Neumann equation (4), but a Lindblad equation. The predictions of the theory differ from those of Ψ-BM and *W*-BM.[5] Advocates of the open-system dynamics should be understood as committing not just to the theoretical possibility of such open-system dynamics, but also assigning a non-zero epistemic probability that it is the correct description of the fundamental law of the universe. Otherwise, it will be epistemically equivalent to holding DMR_C^U.

Thus understood, DMR_O^U is distinct from DMR_C^U.[6] While DMR_C^U can be shown to be empirically equivalent to WFR, because the fundamental dynamics can be of the same type, advocates of DMR_O^U should assign a non-zero probability to predictions that diverge from standard formulations of realist quantum theories. This provides an epistemic argument in favour of DMR_C^U and against DMR_O^U. Insofar as all data are in full agreement with quantum mechanics, DMR_O^U makes the data more surprising than DMR_C^U. Schematically, we may represent the situation as follows:

$$\mathrm{Prob}(E_{QM} \mid \mathrm{DMR}_C^U) > \mathrm{Prob}(E_{QM} \mid \mathrm{DMR}_O^U) \tag{31}$$

and it is rationally permissible to have the following prior assignment:

$$\mathrm{Prob}(\mathrm{DMR}_C^U) = \mathrm{Prob}(\mathrm{DMR}_O^U) \tag{32}$$

We are also allowed to assign a higher prior to DMR_C^U than to DMR_O^U. The reason we are not rationally required to assign an equal or higher prior to the latter, even though every model of DMR_C^U will correspond to a model of DMR_O^U but not vice versa, is because they are distinct hypotheses about the fundamental laws (Chen 2023c; 2024a). Hence, by Bayes's rule, the posterior probability of DMR_C^U, given available data, is higher than that of DMR_O^U.

In other words, DMR_C^U is the conservative and empirically equivalent extension of WFR, so that they are confirmed and disconfirmed to the same degree by empirical evidence, while DMR_O^U takes on extra epistemic risks. If the empirical evidence had been different and if we had seen deviations from quantum mechanics, we might have possessed empirical evidence to favour DMR_O^U over DMR_C^U. But insofar as the empirical evidence is compatible with DMR_C^U and WFR and incompatible with certain versions of DMR_O^U, the opposite is the case.

This does not mean it is not pragmatically fruitful to consider DMR_O^U. Clarifying such a framework can be useful for developing possible quantum gravity or post-quantum theories. But insofar as no empirical evidence deviates from quantum predictions, we do not have an epistemic reason to go beyond DMR_C^U.

14.7 Frequently Asked Questions

In this section, I answer six sets of frequently asked questions regarding DMR, by pointing to relevant discussions in earlier sections and elsewhere in the literature.

14.7.1 Intelligibility

How is DMR an intelligible thesis? Density matrices, by definition, are about partial information or epistemic ignorance. It makes no sense to be a realist about the density matrix.

Answer: Density matrices can play different roles in quantum theory. While some of them encode partial information or epistemic ignorance, it does not follow that all density matrices play such roles. The universal quantum state can be a fundamental mixed state, not arising from entanglement or statistical mixture, as shown in §14.3–14.4.

14.7.2 Familiarity

DMR is unfamiliar. No textbooks in quantum mechanics introduce density matrices as fundamental features of the universe. Why should we consider DMR?

Answer: The lack of familiarity of DMR is not an epistemic reason for taking it less seriously. We can come to a better understanding of DMR by thinking through the definition, examples, interpretations, possible generalisations, and its relation to WFR. Insofar as we do not have decisive reasons against DMR, we should consider DMR as a serious alternative to WFR. In §14.7.6, I also mention some reasons to favour DMR over WFR.

14.7.3 Purification

Even if DMR is intelligible and there may be reasons to consider it, we can always purify the mixed state by adding a factor in a larger Hilbert space. Why can't we just focus on WFR?

Answer: From the perspective of WFR, that is indeed possible. However, absent any reason to think WFR has to be true, DMR is on the table. From the perspective of DMR, the purification is not always physical. The universal state may just be a fundamental mixed state that does not arise from entanglement from the outside, as there is no physical degree of freedom that it is entangled with.

14.7.4 Physical Equivalence

If DMR and WFR are empirically equivalent, are they not the same theory? How are they physically or ontologically distinct?

Answer: They are physically distinct theories, as they prescribe different physical possibilities. Are fundamental universal mixed states physically possible? Yes on DMR but No on WFR. Empirical equivalence does not entail physical or ontological equivalence. Even though DMR and WFR can yield the same predictions, they provide different explanations for how such phenomena arise. For example, a universal mixed state can give rise to particle trajectories, quantum state collapses, and multiverse evolutions that are ontologically different from those arising from a universal pure state.

14.7.5 Occam's Razor

Doesn't Occam's razor favour WFR over DMR? Isn't DMR more complicated than WFR?

Answer: The wave-function realist theories and density-matrix realist theories discussed in §3 employ very similar laws. For example, (4), (5), (10) are no more complex than their wave-function counterparts (1), (2), and (7). Regarding state space structure, both wave functions and density matrices live in the Hilbert space. One may insist that, intuitively speaking, there are more density matrices than wave functions, since every pure state corresponds to a one-dimensional projection density matrix but not every mixed state has a corresponding wave function. Hence, one might make a modal version of Occam's razor by objecting that WFR is more modally parsimonious. The modal razor might work when considering WFR and DMR without any additional laws. However, the argument can work against WFR, because some versions of DMR are more modally parsimonious than versions of WFR. For example, the Wentaculus (a version of DMR) is compatible with exactly one initial quantum state but the quantum Mentaculus (a version of WFR) is compatible with infinitely many (see Chen 2024c).

14.7.6 Theoretical Payoffs

What are the theoretical payoffs of DMR?

Answer: We should take DMR at least as seriously as WFR, regardless of the theoretical payoffs of the former. Since DMR is a clear and intelligible alternative to WFR, and since it is no more complicated than WFR, both of them should be on the table. That being the case, there are many theoretical payoffs of DMR. I find the following most compelling, but it is by no means an exhaustive list:

- *Nature of the quantum state.* DMR provides a new argument for the nomological interpretation of the quantum state. On the nomological interpretation, the universal quantum state is regarded as a nomic object like the Hamiltonian telling objects how to move. However, the exact initial wave function may not be sufficiently simple to be a law. DMR offers a new possibility (Chen 2018, Chen 2022a, Chen 2024c). For example, in the Wentaculus theories, we can stipulate that the initial quantum state is the normalised projection onto the Past Hypothesis subspace, one that is no more complex than the Past Hypothesis itself. Insofar as PH can be simply specified, the initial density matrix can be too.
- *Strong Determinism.* DMR can potentially satisfy our explanatory desire to have all physical facts explained by simple laws. Determinism provides conditional explanations of later states in terms of earlier ones. But that does not explain the initial state of the universe, which requires an additional postulate. With Everettian Wentaculus, the initial density matrix is pinned down by a law, and the multiverse evolution is too (Chen 2024b, Chen 2023b).
- *Statistical Mechanical Probability.* Insofar as we have a fundamental density matrix in DMR that can mimic the statistical density matrix over the underlying pure state of WFR, we can regard the statistical-mechanical probability distribution as being subsumed under the fundamental density matrix and becoming part of the quantum-mechanical probability (Chen 2020).
- *Theoretical Unity.* In a universe with a fundamental wave function, most subsystems may be in mixed states but the universal one is pure. This is not the case in a universe with a fundamental mixed-state density matrix, where typical subsystems and the universe are all in mixed states (Chen 2018).
- *Nomic Exactness.* DMR provides a new route to eliminate fundamental nomic vagueness, i.e., vagueness in the fundamental laws of nature (Chen 2022b).

14.8 Conclusion

DMR is a clear and intelligible thesis. With suitable choices of states and dynamics, DMR is empirically equivalent to WFR. Since they are compatible with different sets of states, they cannot both be correct. It is therefore an interesting case study for scientific realism. In a quantum universe, we face a substantive and interesting choice regarding what kind of quantum theory we should accept, even after choosing our favourite interpretation of quantum theory. Empirical evidence underdetermines between Ψ-theories and W-theories, so our decision needs to rely on theoretical virtues (super-empirical considerations). In a quantum universe, a satisfactory defense of scientific realism must confront this issue and justify what reasons (if any) we have to favour one over the other.

Endnotes

1. The formulation of DMR in that paper seems to require that the universal density matrix be impure, but I intended to allow the possibility that it is pure.
2. It was raised as an objection to W-BM by Albert (2023), but properly understood it is a feature and not a bug of W-BM.
3. It is true that we often know more about the subsystem state than that it is reduced from a typical pure state of the present macrostate. The argument can be modified to accommodate such information. First, restrict ourselves from the present macrostate $\mathscr{H}_M$ to those wave functions whose earlier histories are compatible with a Past Hypothesis $\mathscr{H}_{PH}$. This gives us a smaller subspace $\mathscr{H}_{M'}$. Next, project the wave functions onto the subspace corresponding to the present 'branch' of the wave function, say, with humans existing. This yields another subspace $\mathscr{H}_{M''}$. How the projection is physically implemented depends on the interpretation of quantum mechanics. (In Bohmian theories, we can obtain a much more detailed conditional density matrix of the subsystem, using the environmental configuration located outside the subsystem of interest.) Once we focus on $\mathscr{H}_{M''}$, the argument from canonical typicality applies again.
4. For a similar argument, see Wallace (2016).
5. This theory is suggested but not endorsed in (Allori et al. 2013, sect. 4.3). Thanks to Roderich Tumulka for discussions about this point.
6. DMR_C^U includes W-GRW theories, but their fundamental laws differ from the master equations of open systems, such as the Lindblad equation (see Goldstein et al. 2012, sect. 6.3).

Acknowledgements

For helpful discussions, I am grateful to David Albert, Craig Callender, Eugene Chua, Mike Cuffaro, Sheldon Goldstein, Katie Robertson, Stephan Hartmann, and Roderich Tumulka. I thank the editors and two anonymous reviewers for valuable feedback. This project is supported by an Academic Senate Grant from University of California, San Diego and Grant 62210 from the John Templeton Foundation. The opinions expressed here are those of the author and do not necessarily reflect the views of the John Templeton Foundation.

Bibliography

Albert, D. Z. (1996), 'Elementary Quantum Metaphysics', in J. T. Cushing, A. Fine, and S. Goldstein (eds.), *Bohmian Mechanics and Quantum Theory: An Appraisal* (Dordrecht: Kluwer [Academic Publishers]), 277–84.

Albert, D. Z. (2023), *A Guess at the Riddle: Essays on the Physical Underpinnings of Quantum Mechanics* (Harvard University Press).

Allori, V., Goldstein, S., Tumulka, R., and Zanghì, N. (2013), 'Predictions and Primitive Ontology in Quantum Foundations: A Study of Examples', in *The British Journal for the Philosophy of Science*, 65: 323–52.

Barvinsky, A., and Kamenshchik, A. Y. (2006), 'Cosmological Landscape from Nothing: Some Like it Hot', in *Journal of Cosmology and Astroparticle Physics* 2006: 014.

Chen, E. K. (2018), 'Quantum Mechanics in a Time-Asymmetric Universe: On the Nature of the Initial Quantum State', in *The British Journal for the Philosophy of Science* 72: 1155–83.

Chen, E. K. (2019a), 'Quantum States of a Time-Asymmetric Universe: Wave Function, Density Matrix, and Empirical Equivalence'. Master's Thesis, Department of Mathematics, Rutgers University, New Brunswick. arXiv:1901.08053.
Chen, E. K. (2019b), 'Realism About the Wave Function', in *Philosophy Compass* 14: e12611.
Chen, E. K. (2020), 'Time's Arrow in a Quantum Universe: On the Status of Statistical Mechanical Probabilities', in V. Allori (ed.), *Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature* (Singapore: World Scientific), 479–515.
Chen, E. K. (2022a), 'From Time Asymmetry to Quantum Entanglement: The Humean Unification', in *Noûs* 56: 227–55.
Chen, E. K. (2022b), 'Fundamental Nomic Vagueness', in *The Philosophical Review*, 131: 1–49.
Chen, E. K. (2023a), 'The Past Hypothesis and the Nature of Physical Laws', in B. Loewer, B. Weslake, and E. Winsberg (eds.), *The Probability Map of the Universe: Essays on David Albert's* Time and Chance (Harvard University Press).
Chen, E. K. (2023b), 'The Preordained Quantum Universe', in *Nature* 624: 513–5.
Chen, E. K. (2023c), 'The Simplicity of Physical Laws'. arXiv:2210.08143.
Chen, E. K. (2024a), *Laws of Physics* (Cambridge University Press).
Chen, E. K. (2024b), 'Strong Determinism', in *Philosophers' Imprint* 24:10, 1–24.
Chen, E. K. (2024c), 'The Wentaculus: Density Matrix Realism Meets the Arrow of Time', in A. Bassi, S. Goldstein, R. Tumulka, and N. Zanghì (eds.), *Physics and the Nature of Reality: Essays in Memory of Detlef Dürr* (Springer).
Chen, E. K., and Tumulka, R. (2022), 'Uniform Probability Distribution Over All Density Matrices', in *Quantum Studies: Mathematics and Foundations* 9: 225–33.
Chen, E. K., and Tumulka, R. (2024), 'Typical Quantum States of the Universe are Observationally Indistinguishable'. arXiv:2410.16860.
Chua, E. Y. S., and Chen, E. K. (2025), Decoherence, Branching, and the Born Rule in a Mixed-State Everettian Multiverse. *Synthese* 205, 160.
Connes, A., and Rovelli, C. (1994), 'Von Neumann Algebra Automorphisms and Time-Thermodynamics Relation in Generally Covariant Quantum Theories', in *Classical and Quantum Gravity* 11: 2899.
Cuffaro, M. E., and Hartmann, S. (2021), 'The Open Systems View'. arXiv:2112.11095.
Dürr, D., Goldstein, S., Tumulka, R., and Zanghì, N. (2005), 'On the Role of Density Matrices in Bohmian Mechanics', in *Foundations of Physics* 35: 449–67.
Dürr, D., Goldstein, S., and Zanghì, N. (1992), 'Quantum Equilibrium and the Origin of Absolute Uncertainty', in *Journal of Statistical Physics* 67: 843–907.
Gemmer, J., and Mahler, G. (2003), 'Distribution of Local Entropy in the Hilbert Space of Bi-Partite Quantum Systems: Origin of Jaynes' Principle', in *The European Physical Journal B – Condensed Matter and Complex Systems* 31: 249–57.
Goldstein, S., Lebowitz, J. L., Tumulka, R., and Zanghì, N. (2006), 'Canonical Typicality', in *Physical Review Letters* 96: 050403.
Goldstein, S., Tumulka, R., and Zanghì, N. (2012), 'The Quantum Formalism and the GRW Formalism', in *Journal of Statistical Physics* 149: 142–201.
Hartle, J. B., and Hawking, S. W. (1983), 'Wave Function of the Universe', in *Physical Review D* 28: 2960–75.
Hawking, S. W. (1976), 'Breakdown of Predictability in Gravitational Collapse', in *Physical Review D* 14: 2460.
Hawking, S. W. (1982), 'The Unpredictability of Quantum Gravity', in *Communications in Mathematical Physics* 87: 395–415.

Maroney, O. (2005), 'The Density Matrix in the de Broglie–Bohm Approach', in *Foundations of Physics* 35: 493–510.

McCoy, C. (2020), 'An Alternative Interpretation of Statistical Mechanics', in *Erkenntnis* 85: 1–21.

Ney, A. (2021), *The World in the Wave Function: A Metaphysics for Quantum Physics* (Oxford University Press).

Page, D. N. (1986), 'Density Matrix of the Universe', in *Physical Review D* 34: 2267.

Page, D. N. (2008), 'No-Bang Quantum State of the Cosmos', in *Classical and Quantum Gravity* 25: 154011.

Popescu, S., Short, A. J., and Winter, A. (2006), 'Entanglement and the Foundations of Statistical Mechanics', in *Nature Physics* 2: 754–8.

Robertson, K. (2022), 'In Search of the Holy Grail: How to Reduce the Second Law of Thermodynamics', in *The British Journal for the Philosophy of Science* 73: 987–1020.

Teufel, S., Tumulka, R., and Vogel, C. (2025). 'Canonical Typicality for Other Ensembles than Micro-Canonical', in *Annales Henri Poincaré* 26: 1477–1518

Tumulka, R. (2010), 'Bohmian Mechanics at Space–Time Singularities: II. Spacelike Singularities', in *General Relativity and Gravitation* 42: 303–46.

Wallace, D. (2012), *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation* (Oxford: Oxford University Press).

Wallace, D. (2016), 'Probability and Irreversibility in Modern Statistical Mechanics: Classical and Quantum', forthcoming in D. Bedingham, O. Maroney, and C. Timpson (eds.), *Quantum Foundations of Statistical Mechanics* (Oxford University Press). arXiv:2104.11223v1.

15
Conservation Laws in the Many-Worlds Interpretation of Quantum Mechanics

Lev Vaidman

15.1 Introduction

We are witnessing rapid progress in technology that allows more and more precise measurements, but until today we have not seen a deviation from the predictions of quantum theory. These predictions are made in the standard framework based on analyses of closed systems. Performing experiments on isolated systems provides the expected results. Considering larger and larger systems we extrapolate our understanding of smaller systems and, in the framework of the closed systems approach, we may suggest that the whole universe is a unitarily evolving closed system. This approach leads to the many-worlds interpretation (MWI) which, in my view, is the best available interpretation of quantum mechanics, as it avoids action at a distance and randomness at the level of physics which describes all worlds together (Vaidman 2021).

Although I strongly believe in the existence of a wavefunction of the universe, we do not have direct access to it. We, in our world, know very little about the wavefunction of the universe, thus, the programme of obtaining our observed classical world as emerging from the universal wavefunction (Wallace 2012) has a small feasibility of success. We have access to our world and some information about recently created parallel worlds, but all these worlds contain only a tiny part of the information encoded in the wavefunction of the universe. Although physical laws describe all worlds together, what is relevant for us is the wavefunction of our world. It includes records of our world in the past, and analysis of these records is our scientific task. By observing these records, including descriptions of preparations of quantum states, we can reconstruct the time-dependent wavefunction of our world in the past. (In my understanding, we cannot discuss a single world in which we will be in the future: our world now is related to numerous future worlds.) The wavefunction of my world in the past does not evolve in a unitary way. At every quantum measurement that does not have a single possible outcome, an uncontrollable jump takes place. The description in the world of MWI is exactly the same as in a world of the universe with von Neumann collapses

Lev Vaidman, *Conservation Laws in the Many-Worlds Interpretation of Quantum Mechanics*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press.
© Oxford University Press (2026). DOI: 10.1093/9780198929277.003.0016

at measurements (with the same ambiguity of the exact moment when the measurement obtains a definite outcome). This seems very much like a description of a world in the framework of open systems in which we get uncontrollable disturbances from the environment.

One of the important differences between open and closed system approaches is that we get uncontrollable kicks from the environment leading to failures of some conservation laws which hold for closed systems. When the Hamiltonian of physical interactions has a symmetry, Noether's theorem leads to a conserved quantity. The quantum state of the interacting systems is usually a superposition or mixture of different values of this quantity. Without measurements, this distribution is constant over time. In the standard approach to quantum measurement, if a system is in a superposition of different values of the conserved quantity, its measurement ends up with a single value. In general, the measurement interaction does not change the eigenvalues, but we end up with a singular distribution. So, according to the standard analysis, the distribution changes. Only when we consider the whole ensemble is the distribution the same as represented by the quantum state, and this is the counterpart of the conservation law for situations with measurements. Since in every individual measurement event the expectation value of the conserved quantity is changed, it mimics the open-system situation (Cuffaro and Hartmann 2023).

More close analysis shows some differences: an external system kicks and thus shifts the distribution, or if the interaction with external systems has an uncertainty, it widens the distribution. The measurement-induced change in the distribution of conserved quantities behaves differently. This is what I want to discuss here.

15.2 Evolution of the Distribution of a Conserved Quantity Within a World of the MWI

In this section, I shall prove a general feature of the evolution of conserved quantities within worlds of the MWI. Although I am pretty sure that the result is technically correct, its relevance for our world is far from clear, as will be discussed in the sections which follow.

Let us define our setup. We consider a universe which has symmetry leading to a conserved quantity. For simplicity we assume that it is a translation symmetry in flat space (no gravity, cosmology, etc. are considered). The whole universe which includes all worlds is described by a highly entangled wavefunction $|\Psi\rangle$ that evolves in a unitary way. Let us assume that the universe at the beginning has a particular distribution of momentum $f_{\text{uni}}(\vec{P})$. The density function f is non-negative and normalised: $\int f_{\text{uni}}(\vec{P})d\vec{P} = 1$. The translation symmetry ensures that this distribution is constant in time.

Let us now consider a quantum measurement of some observable O having eigenvalues $\{o_i\}$. The measurement is specified by a set of possible outcomes with a macroscopic pointer that shows the results $\{o_i\}$ in different worlds. In such a situation, the pointer's states are orthogonal, or more generally, the pointer's readings correspond to spaces of quantum states which do not overlap. For a world with an outcome o_i, the momentum distribution is $f_i(\vec{P})$. Let us signify the probability of an outcome o_i by p_i. (In my take on MWI (Vaidman 2021), there is only an illusion of probability, the world splits into multiple worlds with 'measures of existence' p_i; see the definition of this concept in Vaidman (1998).) Probabilities (measures of existence) are defined by the wavefunction $|\Psi\rangle$. Due to the orthogonality of the spaces corresponding to different outcomes, we have the following relation:

$$\sum_i p_i f_i(\vec{P}) = f_{\text{uni}}(\vec{P}). \tag{1}$$

This equation places constraints on what might be the distribution of momentum in a world. The support of $f_i(\vec{P})$ must not be larger than the support of $f_{\text{uni}}(\vec{P})$. Since interference of worlds is not feasible (without the super-technology of Wigner who can perform measurements on his friend), more measurements cannot increase the support of the momentum distribution in a world, so, when more and more measurements are performed, one might expect that the support becomes smaller and smaller, unless there are some special reasons (the possibilities for which will be discussed in sections below) that prevent this change.

By definition, we live in one world. It is hard to estimate how many measurements have been performed in our world. Today, there are many actually working quantum devices, and apparently some natural phenomena also play the role of quantum measurements. (The frequency of such natural cases seems to be a difficult physical question.) Clearly, Stern–Gerlach experiments are quantum measurements; Geiger counters and bubble chambers observe quantum events. Thus, apparently, we should expect a non-conservation of conserved quantities similar to what is happening in open systems.

The earliest experiment exhibiting quantum randomness that I am aware of is a spinthariscope built by William Crookes in 1903 for the observation of single radiation events (see Kolar and den Hollander 2004). Not that I claim that no quantum detection events that split the world have happened before; supernova explosions apparently are also quantum events with macroscopic records.

Even if we do not have clear knowledge about splitting in the past, multiple usage of radiation counters, single-particle detectors, and quantum random-number generators tell us that there are now many, many worlds. If support of distributions of conserved quantities indeed decreases in newly created worlds, we should expect very narrow distributions of conserved quantities in our world. However, when the support becomes very small, it will not be changing anymore,

so there will be a tendency for the future worlds to be closed systems, at least from the point of view of conservation laws.

15.3 Collins–Popescu Conjecture

What can invalidate the conclusions of the previous section? One might think that the measurement device will 'give' momentum to the system so that the total momentum of the system and the measurement device will not be changed. However, we do not see it from the interaction Hamiltonian of the canonical von Neumann measurement, say of the z component of the momentum of the system, $H_{\text{int}} = g(t)P_z P_{MD}$. I also do not see any reason why the momentum of the measuring device, P_{MD}, should be of the order of the momentum of the system to balance the change in momentum. Carroll and Lodman (2021) considered an analogous question regarding energy conservation, providing a detailed analysis of a particular example. Their calculations show that energy exchange with the measuring device cannot explain the energy change of the system.

Another interesting proposal to avoid the effect of measurements on conservation laws is the Collins–Popescu conjecture. Collins and Popescu (2025) argued that the conservation laws are satisfied for every outcome of quantum measurements (without relying on the dynamics of the measuring devices) and not just on average for an ensemble of measurements. At first glance, it seems that the conjecture cannot be true. Almost all variables are not in their eigenstates, including variables of the systems like the momentum, for which we have the conservation law of total momentum. When we measure the momentum of a system which has an uncertainty, clearly its distribution changes, and then the distribution of the total momentum changes. (We accept that the measuring device does not compensate for the change.) What might save the conservation law is that initially the system was entangled with other systems such that the final distribution of the total momentum is the same for any outcome of the measurement. This entanglement tells us that initially the system does not have a pure state, but a mixed state. We, however, frequently observe interference of quantum systems as if they have pure states. Collins and Popescu realised that for a wide enough distribution of momentum of the external system, the required entanglement can be vanishingly small, so we will not be able to see the difference between a slightly mixed state and the pure state. In fact, this situation does not look so special when we analyse the interaction of systems in standard quantum experiments, as I will demonstrate in the following example.

In quantum optics experiments photons coming to a beam splitter end up in a superposition of two orthogonal states: passed and reflected. The photon passing through does not change the state of the beam splitter. The reflected photon,

on the other hand, transfers a finite amount of its momentum to the beamsplitter. However, this transfer leads only to a tiny amount of entanglement, because the uncertainty in the momentum of the beamsplitter in a properly working interferometer has to be much larger than the absolute value of the momentum of the photons in the beam. This is due to the uncertainty principle and the requirement that the beam splitter has to be localised with precision much better than the wave length of the photons. When the photon bounces mirrors inside the interferometer, it also transfers momentum. However, the mirrors also must be well localised, so the momentum uncertainty of the mirrors is much larger than the momentum transfer. This leads to a tiny amount of entanglement which only negligibly spoils the interference. When finally the photon is detected in one of the output ports, the world splits, but the total momentum distribution before splitting is the same as in every one of the obtained worlds: the world with the detection and the world with the negative result of the experiment in which the photon leaves the interferometer via the other port(s). Thus, looking at the history of a particular world, we do not see any change in the distribution of the total momentum.

Apparently, the experiment proposed by Carroll and Lodman (2021) can also be explained by the Collins–Popescu suggestion. Carroll and Lodman consider a protocol with the first step: 'Put a primary system 1 into a known quantum state that is a superposition of energy eigenstates'. In their example, the primary system is a spin in a magnetic field. However, they did not explain in detail how to arrange the spin in a magnetic field without entanglement. I am not convinced by their remark: 'There will be a back-reaction of the magnetic moment on the magnetic field, but that is subdominant to the interaction energy we consider here.' If the magnet does not get entangled with the spin, the spatial wavefunction of the particle will.

Another interesting case is the superoscillations (Aharonov et al. 1991, Popescu 1991, Berry 1994). In a rare case, when a particle is found in a low-amplitude region, its momentum is apparently outside the initial distribution (Aharonov et al. 2021, Aharonov et al. 2023). The apparent paradox of non-conservation can be explained by claiming that the initial superoscillation state is not pure but entangled with some external system with wide distribution. It might be true, but it does not naturally follow from the analysis of the experiment that can prepare the superoscillation example. To prepare the required state we can perform a 'spoiled' weak measurement procedure with a pre- and postselection of a particle with a large spin such that in a weak regime we get an anomalously large value (Berry and Shukla 2011). Consider a Stern–Gerlach experiment similar to the one in which the value 100 is obtained for the z component of a spin-$\frac{1}{2}$ particle (Aharonov et al. 1988). Now, the value $100N$ is obtained for the spin-$\frac{N}{2}$ particle. This corresponds to an anomalously large kick of the particle in the z direction. 'Spoiling' means that our

original spatial wavefunction of the particle had a wide spread (instead of the narrow spread required for weak measurements) in the z direction. The interaction Hamiltonian

$$H = g(t)s_z z, \tag{2}$$

provides a weak coupling if we select a small region around $z = 0$, only then the appropriate postselection of the spin state leads to an anomalous kick. With a wide initial z distribution, the spatial wavefunction after the spin postselection has a familiar superoscillation form with a tiny amplitude in the centre, where the spin state was not significantly changed, so it was almost orthogonal to the spin postselection measurement. The exponential increase in amplitude is explained by rotation of the initial spin by the interaction Hamiltonian (2), making the scalar product of local spin at a large distance from $z = 0$ much bigger.

A paradox with a similar flavour was presented by Dicke (1981). A photon beam passes through part of the location of the wave packet of a particle in a ground state without scattering, thus without changing its energy, while the particle increases the energy since it is no longer in the ground state. In a follow-up paper, Dicke (1986) claimed that we can observe energy conservation only on an ensemble. See also the recent related discussion in Sen (2023).

15.4 The Issue of the Reference Frame

The interaction Hamiltonian of the Stern–Gerlach device (2) changes the momentum of the pointer (z component of the momentum of the particle), so the conservation of the total momentum should not be expected. In reality, we do expect symmetry under translation, so we do expect conservation of total momentum. What really appears in the measurement interaction Hamiltonian is

$$H = g(t)S_z(z - z_{MD}). \tag{3}$$

We never measure absolute variables, always relative variables: the system relative to the measuring device (see Yang (2020) and references therein). The conservation law of momentum and the Third Law of Newton are closely related, so measurement interaction does not change the total momentum. The total momentum commutes with $z - z_{MD}$. This observation might suggest that measurements and conservation laws are not connected. It also questions the significance of the Araki–Yanase theorem (Araki and Yanase 1960), as well as its recent generalisation to continuous variables (Kuramochi and Tajima 2023).

However, the separability of the global variables and relative variables is true only if there is no entanglement between the relative variables and the global

variables. This assumption is closely connected to the Collins–Popescu conjecture. Indeed, the basic requirement for measurement is that measuring devices are not entangled with measured systems prior to measurement. If the measured system is in a pure state which is a superposition of eigenstates of momentum, then we obtain entanglement between relative variables (of the system and the measuring device) and the global momentum: observation of smaller momentum corresponds to smaller total momentum, while the observation of larger momentum of the system corresponds to larger total momentum. The entanglement of the system with the 'preparer' allows Collins and Popescu to avoid entanglement of the global variables and relative (system relative to measuring device) variables. Under the assumption that the momentum distribution of the measuring device is not changed, the Collins–Popescu mechanism leads to the same distribution of total momentum for observation of smaller or larger momentum of the system.

An even stronger assumption is that the universe is in an eigenstate of the total momentum. Apparently, it is consistent with our experience of the world when we observe only relative positions. Motivation for such a constraint might appear in the cosmological analysis of the beginning of the universe which, as far as I know, has no established results yet. Anyway, the construction of the relative variables seems to be simpler than the Page–Wootters scheme (Page and Wootters 1983) required for a universe in the eigenstate of energy, for which a variable which replaces time has to be invented.

15.5 Conclusions

In Section 15.2 I presented an (as far as I know new) simple strong formal result which puts constraints on the distribution of conserved quantities within the worlds of the MWI. In every world of the MWI quantum measurements do not increase the support of distributions of conserved quantities. Thus, measurements tend to decrease the uncertainty of the global conserved quantities. This regime has some similarity with the non-conservation of conserved quantities of open systems. In the long run, however, there might be saturation to very small uncertainties and non-conservation will become negligible.

In Section 15.4 I argued that since all measurements are measurements of relative variables which commute with global conserved variables, the effect of measurements on conservation laws is not immediately expected. However, the connection is possible if for some (cosmological?) reason we do have initial entanglement between global and relative variables. So, the mechanism described in Section 15.2 might change distribution of conserved quantities in separate worlds of the MWI keeping the conservation law for the universe which includes all the worlds.

Section 15.3 is devoted to the Collins–Popescu conjecture, which also describes initial entanglement. This is a special type of tiny entanglement between an observed system and an external system, 'preparer', which has large uncertainties of the conserved variable. This entanglement cancels the entanglement between relative system—measuring device variables and global conserved variables, and thus avoids the connection between measurements and conservation laws. The Collins–Popescu special type of entanglement is obtained naturally in some typical experimental situations. More general analysis is needed to shed light on the actual scenario that corresponds to the initial state of the universe.

Living in a particular world of many worlds (which can be properly analysed only in the past) we witness a non-unitary evolution (an illusion of collapse) which is not described by the standard closed system approach. The question of conservation laws in this world is not given a definitive answer. For example, the total momentum of the universe might correspond to a particular value (relative to some quantum frame of reference), it can be a distribution constant in time, or its uncertainty can decrease with more measurements to be performed. In all cases the conservation laws differ from the standard case of the open system approach in which we do not expect exact conservation laws to hold.

Even if the conservation laws hold in the splitting process of the MWI, the evolution of the quantum state of a world is not unitary, so a world cannot be considered as a closed system. It is important to understand this, since it might be that some paradoxes can be resolved by appreciating this fact. In discussing the 'unitarity puzzle' appearing in the evaporation process of a black hole predicted by Hawking (1976), we can put an arbitrary amount of information into a black hole, but when it evaporates, we get just thermal radiation, so the information is lost, and unitarity fails (see Giddings 2006). A simple resolution might be as follows. It is usually assumed that we end up with one world with thermal radiation. However, evaporation of a black hole is a quantum process, so it happens at different times. Thus, we get a superposition of thermal radiations created at different times. Relative phases between these terms are capable of containing the information that was fed into the black hole. See Bao et al. (2018) and Wang et al. (2023) and references therein for recent alternative analyses.

Acknowledgements

This work has been supported in part by the Israel Science Foundation Grant No. 2689/23 and the Bristol Benjamin Meaker Distinguished Visiting Professorships scheme.

Bibliography

Aharonov, Y., Albert, D. Z., and Vaidman, L. (1988), 'How the Result of a Measurement of a Component of the Spin of a Spin-1/2 Particle Can Turn Out to be 100', in *Physical Review Letters* 60: 1351.

Aharonov, Y., Popescu, S., and Rohrlich, D. (1991), *Preprint TAUP 847-90* (Tel Aviv University).

Aharonov, Y., Popescu, S., and Rohrlich, D. (2021), 'On Conservation Laws in Quantum Mechanics', in *Proceedings of the National Academy of Sciences* 118: e1921529118.

Aharonov, Y., Popescu, S., and Rohrlich, D. (2023), 'Conservation Laws and the Foundations of Quantum Mechanics', in *Proceedings of the National Academy of Sciences* 120: e2220810120.

Araki, H., and Yanase, M. (1960), 'Measurement of Quantum Mechanical Operators', in *Physical Review* 120: 622.

Bao, N., Carroll, S. M., Chatwin-Davies, A., Pollack, J., and Remmen, G. N. (2018), 'Branches of the Black Hole Wave Function Need Not Contain Firewalls', in *Physical Review D* 97: 126014.

Berry, M. (1994), 'Faster than Fourier', in J. S. Anandan and J. L. Safko (eds.), *Quantum Coherence and Reality*: In Celebration of the 60th Birthday of Yakir Aharonov (World Scientific) 55–65.

Berry, M., and Shukla, P. (2011), 'Pointer Supershifts and Superoscillations in Weak Measurements', in *Journal of Physics A: Mathematical and Theoretical* 45: 015301.

Carroll, S. M., and Lodman, J. (2021), 'Energy Non-Conservation in Quantum Mechanics', in *Foundations of Physics* 51: 83.

Collins, D., and Popescu, S. (2025), 'Conservation Laws for Every Quantum Measurement Outcome', in *Quantum* 9: 1815.

Cuffaro, M. E., and Hartmann, S. (2023), 'The Open Systems View and the Everett Interpretation', in *Quantum Reports* 5: 418–25.

Dicke, R. (1981), 'Interaction-Free Quantum Measurements: A Paradox', in *American Journal of Physics* 49: 925–30.

Dicke, R. (1986), 'On Observing the Absence of an Atom', in *Foundations of Physics* 16: 107–13.

Giddings, S. B. (2006), 'Black Hole Information, Unitarity, and Nonlocality', in *Physical Review D* 74: 106005.

Hawking, S. W. (1976), 'Breakdown of Predictability in Gravitational Collapse', in *Physical Review D* 14: 2460.

Kolar, Z., and den Hollander, W. (2004), '2003: A Centennial of Spinthariscope and Scintillation Counting', in *Applied Radiation and Isotopes* 61: 261–6.

Kuramochi, Y., and Tajima, H. (2023), 'Wigner–Araki–Yanase Theorem for Continuous and Unbounded Conserved Observables', in *Physical Review Letters* 131: 210201.

Page, D. N., and Wootters, W. K. (1983), 'Evolution Without Evolution: Dynamics Described by Stationary Observables', in *Physical Review D* 27: 2885.

Popescu, S. (1991), *Ph.D thesis* (Tel Aviv University).

Sen, R. (2023). 'The Collapse Problem as a Consistency Problem: Is the Quantum Measurement Hypothesis Consistent with Conservation Laws?', in *Annals of Physics* 458: 169462.

Vaidman, L. (1998), 'On Schizophrenic Experiences of the Neutron or Why We Should Believe in the Many-Worlds Interpretation of Quantum Theory', *International Studies in the Philosophy of Science* 12: 245–61.

Vaidman, L. (2021), 'Many-Worlds Interpretation of Quantum Mechanics', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University). Fall 2021 edition. URL https://plato.stanford.edu/archives/fall2021/entries/qm-manyworlds/.

Wallace, D. (2012), *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation* (Oxford University Press).

Wang, X., Zhang, K., and Wang, J. (2023), 'Entanglement Islands, Fire Walls and State Paradox from Quantum Teleportation and Entanglement Swapping', in *Classical and Quantum Gravity* 40: 095012.

Yang, J. M. (2020), 'Switching Quantum Reference Frames for Quantum Measurement', in *Quantum* 4: 283.

16
Quantum Theory: Ideals, Infinities, and Pluralities

Gemma De les Coves

16.1 Overture

Emulating the maxim of enlightened despotism: *Everything for the people but without the people*, the maxim of physics seems to be: *Everything for reality but without reality*. The only case we can solve exactly is the ideal case, which tends to involve an infinity—the thermodynamic limit, zero temperature ($1/\infty$), particles coming from infinity, infinitely long plates … We then master a conceptually delicate leap to the 'real' case, resulting in finite size effects, finite temperature, perturbations … I will argue that the ideal cannot exist in the physical world because it necessarily involves an infinity, and infinity cannot be supported in the world. The ideal, thus, acts as a ghostly pivot (as its existence is excluded from this world) from which we build a bridge to the real case. What kind of bridge is that?

The case that will be the centre of our attention is that of quantum theory, where closed quantum systems (the ideal case) can be used to provide a description of open quantum systems (the real case) via purifications. Quantum theory stands out among other theories in the fact that, in its appeal to the ideal, it need not involve an infinity in the dimension of the system. I will nonetheless argue that it needs to involve an infinity in its description, from which it will follow that closed quantum systems cannot exist in the world.

Inspired by the tension between transcendence and closure, I will then argue that the universe should not be considered as a closed quantum system. The most profound lesson of these considerations is, I believe, the contradictory nature of the notion of an ultimate closure—essentially because to be able to conceive of a unity implies putting a boundary, and that boundary could be transcended, frustrating the claim that the former unit was ultimate. Conceiving of the universe as a closed system would amount to positing an ultimate closure. I will conclude that, if anything, the universe should be conceived of as a plurality.

Before delving into this perspective, a disclaimer: I don't know what is real, how to define reality, or if reality exists (I don't even know what it means to exist).

Gemma De les Coves, *Quantum Theory: Ideals, Infinities, and Pluralities*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press.
© Oxford University Press (2026). DOI: 10.1093/9780198929277.003.0017

But I will pretend I do, in particular I will assume that there exists a physical reality.[1] Throughout this text I will use 'real' as a synonym for 'physically relevant', which should mean that it can be instantiated in the physical world. 'Ideal' will be opposed to 'real'. The choice of words is unfortunate, as the real (in reference of to reality) is in fact the ideal in the sense of the unattainable—what (mysteriously) draws and guides us. Another word of caution: I don't understand the physics of quantum theory, I just know its mathematics. Finally I wish to warn the reader that I will refrain from explaining (highly technical) mathematical concepts and results, while I may explain concepts that readers with a philosophical background may be familiar with. This imbalance reflects my state of knowledge.

Let us start this journey by walking the bridge from the real to the ideal in quantum theory.

16.2 The Real and Ideal in Quantum Theory

As far as I see, ideal quantum systems cannot exist in the physical world for at least three reasons. The first is ontological, the second is epistemic, and the third concerns the world's incapability to instantiate infinity.[2]

The first reason is plain but far-reaching: no system can be completely isolated.

The second reason goes as follows. An ideal quantum system is assumed to be an object that can be defined and described in an isolated manner. So it does not only presuppose an ontic ideal (a perfectly isolated system) but also an epistemic ideal (whose state can be described perfectly). In particular, it posits a radical subject–object split. If such a Cartesian split is generally contested (Gefter 2023), then even more so in the case of quantum theory. At least during a measurement the quantum system must interact with the outside.

In fact, the assumptions of complete isolation and the existence of another system that can describe the first one seem to contradict each other. For if a system were completely isolated, no other system could know anything about it. Namely, if we assume that knowledge supervenes on the state of a physical system and that knowledge acquisition is enabled by a physical process, a state of knowledge should be expressible as a correlation between the physical system on which it supervenes and the state of the other system, and knowledge acquisition as the construction of such a correlation. (This view is more explicit in Fankhauser et al. (2024).) And correlating is only possible by means of an interaction between the two systems, which contradicts the purported total isolation of the system.[3] Note that total isolation is understood as not only in the present but also in the past.

The third reason is that the mathematical representation of ideal quantum systems necessarily involves an infinity, and infinity cannot be supported in the physical world. Making this argument will occupy us for much of this section. Throughout the journey, we will consider the relation of other physical theories

with infinity, mathematics, and computer science. We will also shed a mereological and epistemological light on some of our digressions.

Note that, for the universe itself, the first reason is moot, the second too (if the agent is to be outside the universe), and the third unclear. I will argue in section 16.3 (The Universe as a Plurality) that the universe should not be considered a closed quantum system for other reasons, grounded in the tension between transcendence and closure.

Let us now set the stage.

16.2.1 The Real and Ideal in Quantum Theory

In quantum theory, the ideal corresponds to closed quantum systems, which describe the situation where the physical system is completely isolated from the rest of the world, devoid of any interaction with the rest. Such systems are mathematically described by pure states. I stick to the interpretation by which a pure state, identified with a rank-one density operator, represents perfect knowledge of the system. So for the kinematics of quantum systems, with regard to the dynamics, the ideal corresponds to unitary evolution.

In quantum theory, the real corresponds to open quantum systems, which describe physical systems interacting with the rest. Mathematically, they are described by mixed states. With regard to the dynamics, the real corresponds to quantum channels, that is, completely positive trace-preserving linear maps.

As far as the mathematical formalism goes, the ideal is a special case of the real. In other words, the formalism of open systems is more expressive than that of closed systems. Let me say a few a more things about the mathematical structure of our description of open quantum systems.

16.2.2 Open Quantum Systems, Mathematically

Density operators are given by positive semidefinite matrices,[4] which embody a 'genuinely two-dimensional notion' of positivity. Namely, nonnegativity is unambiguous for a number $x \geq 0$, which can be thought of as a zero-dimensional array. For one-dimensional arrays, the only meaningful notion of nonnegativity are nonnegative vectors, $v = (v_1, \ldots, v_n)$, such that $v_i \geq 0$ for all i. Two-dimensional arrays can be nonnegative in the sense of (a) a nonnegative matrix, which is a matrix whose every entry is nonnegative, or (b) positive semidefinite, which is a genuinely matrix notion. Both notions form a cone, although with very different properties (see, e.g., De les Coves et al. 2023). I wonder if there is a three-dimensional notion of nonnegativity, represented by cubes (a.k.a. rank-3 tensors), or higher-dimensional notions. Why is the two-dimensional notion of

nonnegativity (provided by positive semidefinite matrices) so adequate to describe open quantum systems? Note that the one-dimensional notion is very useful to describe classical theories of nonnegativity, such as probability distributions and stochastic matrices (see, e.g., Horn & Johnson 1985).

Composition plays a central role in the theory of (open) quantum systems, coming to the foreground in the distinction between positive and completely positive maps, which is intimately linked with entanglement. In my eyes, the surprising fact of complete positivity is that it seems to refer to an extrinsic property of the system where the map is acting, by involving an 'innocent bystander'. As a matter of fact, this is just the tip of the iceberg of the rich relations between the composition of systems and positivity. Mathematically, composition is given by the tensor product of vector spaces, resulting in a new vector space, whereas positivity forms a cone, and the tensor product of the cones is not the cone of the tensor products. The cone of positive semidefinite matrices is special in several ways; for example, it is self-dual (De les Coves et al. 2023). If composition is important, so is its converse, namely 'ignoring', mathematically modelled by the partial trace. As we will see below, this operation provides the bridge from the ideal to the real.

Turning to the dynamics of quantum systems, we encounter a contested issue (invisible in the kinematics): whether transformations of open quantum systems should be required to be continuous in time. Surprisingly, most quantum channels do not describe a continuous time evolution; only a tiny minority do, i.e., they can be expressed as e^{tL} where L is Liouvillian of Lindblad form (Wolf & Cirac 2008). An even tinier minority express *reversible* dynamics, namely those of the form e^{tL} where L only has a Hamiltonian part, rendering the evolution unitary. The only quantum channels whose inverse is also a quantum channel are precisely the ideal ones, i.e., unitary channels. Even more, quantum channels may not be divisible once.[5] It follows that, *according to the theory of open quantum systems*, the evolution of a quantum system (whether closed or open) need not be reversible or continuous. I believe I understand why quantum channels describe irreversible dynamics: because they introduce some noise in the system that cannot be undone. Technically, they are non-injective, because of the contraction following the purification. But why should they also be non-continuous? The theory of open quantum systems, as far as the dynamics is concerned, does not posit an infinite divisibility of time, as other physical theories do. In fact, the theory seems to inform us that time need not have a continuous nature.

16.2.3 Resorting to the Ideal in Quantum Theory

Mixed states can be described without appealing to pure states, namely as positive semidefinite matrices, which are objects A such that $v^*Av \geq 0$ for all v. By the spectral theorem, they can also be expressed as a convex combination of projections

to pure states, $A = \sum_i p_i v_i v_i^*$, where this expression can be seen as a special case of Stinespring's dilation theorem. The appeal to the ideal followed by a bridge to the real is transparent: the ideal takes the form of a rank-one density operator followed by a contraction resulting in the convex combination.

Because of the isomorphism between linear maps and elements in a tensor product space, as well as some easy equivalences (known as the Choi–Jamiołkowski isomorphism), similar considerations apply to transformations of quantum states, where the unitary evolution corresponds to the ideal case and completely positive maps to the real case. Completely positive maps can be described directly as linear maps whose extension with the identity of any size[6] is positive, or via the Kraus decomposition, which is grounded on the same transparent appeal to the ideal via Stinespring's dilation. And, yet again, similar considerations apply to measurement operators described by positive operator-valued measurements (POVMs; corresponding to the real case) versus projective measurements (corresponding to the ideal case). All these mathematical objects describing aspects of open quantum systems can be purified, i.e., given in terms of a modified ideal. This is extremely useful, if only conceptually surprising.

Such purifications imply that the description of open systems can be derived from that of closed systems, from which one may be tempted to conclude that closed systems are ontologically more fundamental than open systems, as the latter seem to be a 'special case' of the former. But this conclusion would be mistaken, I believe, for two reasons. One is that closed quantum systems do not exist in the physical world, as I will try to argue below, rendering our fulcrum to the real world truly ghostly.[7] The second is that it embodies a reductionistic mindset, according to which one should focus on the case from which others can be derived, and obtain the description of the latter from the former. Such a mindset has been proven not only narrow but also misleading on many fronts (we will consider emergence on page 301). Taken to the extreme, this view may imply that culture and art are dispensable, as they strive for the singular whereas science strives for the universal, and the singular is in some sense derivable from the universal (De les Coves 2024). Science is said to provide the view from nowhere, but I have a single dirty window from where to observe the world. While I somehow transcend this window with the help of reason, I believe there is value in the singular, precisely because it is unique and irreducible.[8]

Let me end with two remarks. The first is that purifying has its own *mathematical* pitfalls. For example, purifying a mixed state may result in an uncontrolled increase of the cost of the representation, as measured by a rank (De las Cuevas et al. 2013).[9] A further difficulty of purifications, both conceptual and mathematical, is their non-uniqueness. Conceptually, it leads to interpretation ambiguities. By the very nature of the ideal case, we lack an operational criterion to pick one, as we can only access its 'shadow' (its projection or compression to the real case). Mathematically, it renders the search for an optimal purification (in terms of cost) NP-hard.

The second remark concerns quantum magic squares, which are sets of POVMs that are compatible in a certain 'meshed' way. Their description cannot be obtained from the ideal: not every quantum magic square can be purified. Technically, not every quantum magic square can be dilated to a quantum permutation matrix (De las Cuevas et al. 2020). What is the physical relevance of this result? Quantum magic squares should relate to strategies for quantum games (see, e.g., Lupini et al. 2020), where the lack of a dilation has certain mathematical implications. However, the conceptual dimension of this fact still needs to be explored. The result suggests that quantum magic squares ought to be taken as fundamental, instead of quantum permutation matrices. In this case, there is a mathematical advantage of considering (the analogue of) the real case, as quantum magic squares are strictly more expressive than quantum permutation matrices together with their notion of contraction.

16.2.4 The Ideal in Physics Tends to Involve an Infinity

In physics, the ideal tends to involve an infinity. Phase transitions are defined at the thermodynamic limit, $N \to \infty$, where N is the number of particles. Similarly, the zero-temperature case plays a pivotal role, where 0 is seen as the infinitely small, $1/\infty$. (In this case there is even a principle of thermodynamics preventing the ideal from being realisable in the physical world, which does not prevent us from first solving the zero-temperature case and then describing finite temperature as perturbations around it.) Ground states play an analogous role to zero temperature. We also often consider particles coming from infinity (as in the definition of the scattering matrix), infinite plates (say, with a certain charge density), perfectly round things, plane waves … All these idealisations involve an infinity.

If the ideal is represented by an infinity and we contend that the real is represented by a finite number, the relation from the real to the ideal becomes mereological: from a part to the whole. In other words, because the finite is part of the infinite, if the finite represents the real and the infinite the ideal, then the real can be seen as a part of the ideal. Writing $A \lhd B$ to denote that A is a part of B, we find the following relations analogous:

In Physics: (ȸ)
Real $\lhd$ Ideal
Finite $\lhd$ Infinite

The latter infinity tends to be actual infinity in physics and potential infinity in computer science, as we will discuss on page 302.

One remarkable thing about quantum theory is that the ideal need not involve an infinity, at least in the dimension of the quantum system. This follows from Stinespring's dilation theorem together with the Choi–Effros theorem.[10] It results in the following analogies:

In Quantum Theory: (ℏ)
Real ◁ Ideal
Finite ◁ Finite

The finite on the right-hand side is of square dimension compared to that on the left.

Why can we bound the dimension of the ideal quantum system? We can give a mathematical answer: because of the very stringent properties of completely positive linear maps. Is there a deeper, conceptual reason? I don't know. But I want to make the case that the description of the ideal in quantum theory necessarily involves an infinity. First note that the dimension of the Hilbert space can be inferred only after an unbounded number of measurements of identical copies. This involves two infinities: one in the number of copies, and one in the identicality condition, which involves infinite precision. So the fact that the dimension of the Hilbert space of the ideal quantum system is finite does not imply that it be accessible with a finite number of physical operations. Second, an ideal quantum system represents a limiting case, that of complete isolation, i.e., of no interaction whatsoever with the outside. For this reason, its description must involve arbitrary precision: the interaction between the system and the environment must be smaller than *any* given precision. If this mathematical description is to be consistent with that of open quantum systems and every other area of physics I can think of, the statement of 'any given precision' must involve an infinity. In particular, it involves the potential infinity (by division) and, in our usual mathematical formulation, the actual infinity too (see page 301). In contrast, open quantum systems are not associated with such radical views, neither ontic (the system could have a dependence, even if tiny, on the environment) nor epistemic (they allow to represent my uncertainty). As a consequence, their mathematical descriptions need not involve an infinity. For example, a density operator could be given in terms of rationals with a finite precision (see page 301).

Let us now consider infinity's relation with the world (with our finite and fallible minds).

16.2.5 Infinity and the World

Can the physical world support infinity? I want to answer 'no': the world cannot instantiate infinity. To that end, I will partially rely on Aristotle's answer. He

distinguished between potential and actual infinity, and argued that the potential infinite can exist in the physical world but actual infinity cannot (Moore 2019: Ch. 2). Namely, he defended that there is no objection to something's being infinite in the world provided that its infinity is not there all at once.[11]

In modern terms, I take the potential infinity to be the unbounded, and the actual infinity to be (isomorphic to) an infinite set, that is, a set whose cardinality is at least $\aleph_0$. This aligns with Priest's argument that every potential infinity needs the actual infinity to be defined, by the so-called Domain Principle (Priest 2001: Ch. 8). The potential infinity aligns with Hegel's 'false infinity' and the actual infinity to his 'true infinity' (Priest 2001: Ch. 10).

As far as I understand, Aristotle's reasons for accepting the potential infinity's physical existence were grounded in his belief that time was infinite by addition—he assumed that time is unbounded in the past and the future. With today's knowledge, the first seems to be false and the second is unclear. The space of the universe also seems to be finite. Hence I conclude that the potential infinity cannot be supported in time or space.[12] And, if the foregoing conclusion is true, I posit that potential infinity by addition cannot be supported at all in the physical world. With regard to potential infinity by division, either in time or space, I do not have any reason to think they can be instantiated in the physical world. Why should I believe that something can be divided arbitrarily many times?[13] This something could be empty space, or time. Infinity is a construct with such extravagant properties (page 310) that, in the absence of a reason to believe they can be exemplified in the world, I find it is more sober to believe they cannot. This will be crucial to my argument why ideal quantum systems cannot exist (page 304).

Despite Aristotle's conclusion,[14] Greek geometry presupposes infinite space by addition and by division—for example, a line is indefinitely extensible and infinitely divisible. On the other hand, arithmetic presupposes the existence of infinitely many natural numbers, since each natural number has an immediate successor but no natural number is itself infinite. This rationale was then exported to axiomatisation, whose main appeal is its pretension to trap an infinite wealth of information or wisdom in a finite, manageable stock of basic (self-evident) principles. These hopes were capped by Gödel (see page 307).

The foregoing considerations seem to beg the question of the relation between mathematics, more generally, and the world. In contrast to infinity, some numbers—say, the number three—seem to be instantiated in the world, despite the fact that the number three as such does not exist in the physical world. (That 'three' is instantiated in a collection of (three) pencils, fingers, or moons means somehow that we identify 'three' with the equivalence class of all things which are three.) Similarly, relations such as $3 + 2 = 5$ seem to be instantiated in the physical world, as we never observe that three oranges together with two oranges make anything other than five oranges. Geometrical objects such as circles, lines, or points

are only poorly instantiated—probably because their precise instantiation would require supporting infinity.

In my eyes, the relation between mathematics and the world is a particular case of a broader problem, the problem of universals in metaphysics (see, e.g., Loux & Crisp 2017). Using the term (and communicating the meaning) 'three' is not so different from uttering other lexical items (such as 'run', 'scissors', 'sleep', or 'kitchen'). It requires recognising unity in infinite diversity. This seems to be very close to the core of the problem of cognition. It is also related to the emergence of meaning, to which we shall briefly return on page 307 (for broader considerations, see, e.g., Eco 1997).

16.2.6 The Theory of Physics and Infinity

Despite the world's inability to sustain actual or potential infinity (as argued above), physical theories heavily rely on infinity—usually, actual infinity. For example, we describe space, time, energy, or temperature with real numbers. I believe that the potential infinity, both by addition and division, would suffice, as we only need to invoke finitely many multiples or fractions of fundamental units, such as the metre, the kilogram, or the second. Such fractions or multiples should be finite but unbounded, leading to rational multiples of the fundamental units. And each fundamental unit need not involve an infinity, as it is taken as a 'brute fact' stemming from the world which we use as a yardstick (pardon the redundancy) to measure other such worldly facts.

Let me make the same point from a different angle. Most will agree that what is physically accessible (i.e., what is currently measurable) are rational multiples of the fundamental units with a certain precision. Since we presume that precision will improve with time (as we happen to live in a magnificent epoch of constant discovery), it would be absurd and soon obsolete to develop a theory for a specific precision. Granting that, we would want a physical theory that accounts for *any* precision, that is, that aims to describe rational multiples of the units, instead of real (in the sense of $\mathbb{R}$) multiples, which are their completion. Each real is the limit of a sequence (in fact, of many sequences) of rationals, e.g., π, e, $\sqrt{2}$ or literally countless others. It is uncontested that the notion of a limit is *mathematically* very powerful—in analysis, continuity, derivatives, and integrals rely on limits, to cite some. But we should not be dazzled by the bright light of infinities in mathematics and mistake them for something real.

Del Santo & Gisin (2019) have recently challenged the use of real numbers in physics. They find that if classical physics were only to use quantities admitting a finite description, it would be indeterministic. So they regard real numbers as the hidden variables of classical mechanics, in analogy with the search for hidden variables in quantum mechanics.

In quantum theory, we use the reals 'twice' (in an intertwined way), namely the complex.[15] Renou et al. (2021) have recently shown that such an imaginary part is necessary in quantum theory.[16] But why not use *hyperreals*, instead of reals, also with an imaginary part? The hyperreals are an extension of the reals (Goldblatt 1998) which allow one to solve long-standing questions in quantum information, such as the existence of non-trivial tensor-stable positive maps, which are intimately related to the existence of NPT-bound entanglement (Van der Eyden et al. 2022). The only superficial virtue I see in the reals over the hyperreals is that the former are simpler, so may seem preferable by Occam's razor. But the two have the same type of infinity ($\aleph_1$), and I would argue that reals are already 'full of monsters'—for example, essentially no element in $\mathbb{R}$ admits a finite description, and how can we even conceive of something infinite whose only description is itself? More reflections on reaching infinity from the finite can be found in De les Coves et al. (2024).

Physics' artificial relation with infinity comes to the foreground when it aims to connect with computer science, which relies on potential infinity—for example, the computational complexity of a physical problem can only be studied if the problem is discretised.

16.2.7 The Potential Infinity in Computer Science

The theory of computation is based on the distinction between the finite and the potentially infinite. A Turing machine consists of a finite number of transformation rules, where each symbol is an element of a finite alphabet. The crucial attribute of these elements is *finitude*: finite things are combined in a mechanistic way. The only component that is unbounded is the length of the tape. That a Turing machine halts means that it *eventually* halts, that is, after a finite but unbounded time. 'Dually', a formal language L is defined as $L \subseteq \Sigma^*$ where Σ is a finite alphabet and Σ^* the set of concatenations of elements of Σ *of any length*. Σ^* represents the unbounded. If L is finite, its complexity is trivial within the standard theory of computation, because there is a Turing machine whose transition rules are precisely the set of elements of L, so it accepts L in one time step. In fact, a finite-state automaton without loops suffices. I propose referring to such machines as *dull*, because they distort the very idea that an algorithm is a construct that enables us to reach infinity from the finite. Indeed, I see the theory of computation as a systematic study of which infinite sets can be described with a finite number of rules. Note that this concerns the potential infinite by addition, not by division.

Yet how meaningful is the notion of a halting machine *physically*? Not very, I contend. The Turing machine is required to *eventually* halt, there is no bound at, say, 590875 billion steps—just at some finite number. Physically this is ridiculous,

as the requirement of a finite bound (without any further qualification) corresponds to no time scale at all. Some will counter that this is the rationale behind the class P, which requires the machine to halt in polynomial time. I believe they are partially right, but I would add that P relies on two requirements which involve an infinity. The first is asymptotic scaling, by which the length of the string is unbounded. The second is worst-case analysis, by which the hardest of infinitely many instances determines the complexity of the problem. I believe the reason for demanding these conditions is, again, mathematical instead of physical. For example, average-case complexity has an uglier mathematical structure—e.g., the 'harder-than' relation need not be transitive (Wigderson 2019).[17]

It follows from these considerations, together with the physical world's incapability to sustain infinity, that it is perilous, if not flawed, to measure the complexity of physical problems based on the theory of computational complexity. One such pitfall are the undecidability results, e.g., Wolf et al. (2011) and Cubitt et al. (2015). More generally, I am unclear on the importance of computational complexity results for physics (and have personally worked on them for quite some time, e.g., Klingler et al. 2023). Ultimately, we are asking: What relation should the theory of physics have with potential infinity? On page 302 I argued that it may be sensible for physical theories to describe the potentially infinite (not only by addition but also by division).

Both physics and computer science strive for the infinite from the finite, from a certain perspective, in a similar way: local physical interactions can be seen as the *grammar* of physics (Stengele et al. 2023, Reinhart & De las Cuevas 2022). Such pursuit is not exclusive to these domains of enquiry. In the cognitive sphere, Quine argued that it was only because there are infinitely many things that we need to operate with the fundamental notion of a thing at all (Moore 2019: Ch. 9). For this notion is used principally in making generalisations, by specifying, one by one, what each was like. This, again, lies at the heart of the problem of universals in metaphysics.

16.2.8 Epistemologically

What can be better known, the real or the ideal? Mathematically, the ideal over the real—I think this is the very *raison d'être* of the ideal. Together with (ȸ) this seems to imply that:[18]

In Physical Theories (appealing to the Ideal):

Knowledge of the part < Knowledge of the whole

But this was too fast. For, in reductionistic approaches, the parts can be better known than the whole, probably because they do not appeal to ideals:

In Reductionistic Physical Theories:

Knowledge of the part > Knowledge of the whole

Much of condensed matter physics, approximation techniques, etc., are concerned with bridging this gap. A more or less good knowledge of the parts degrades as they are composed, because the description becomes too difficult (there are too many parameters) and we lose control over the whole.

As a matter of fact, condensed matter physics is riddled with emergent phenomena, associated with the appearance of new degrees of freedom which behave differently than the underlying ones.[19] It follows that the knowledge of the whole we are after is essentially independent of the knowledge of the parts. In my eyes, emergence showcases that a theory may be more ontologically fundamental than another, yet have *less* explicatory power—in terms of identifying the relevant degrees of freedom, explaining their behaviour, and predicting.

In quantum theory, the observation that the ideal can be better known than the real, together with (ħ), implies that:

In Quantum Theory:

Knowledge of the part < Knowledge of the whole

The knowledge, or lack thereof, on either side is optimal for maximally entangled states, which describe perfect knowledge of the whole and perfect ignorance of the parts (in fact, this is one way to define these states). Even more, by the purification theorem (page 296), any lack of knowledge of the state of the system can be attributed to it being entangled with an unspecified, so-far-extrinsic environment, so that together they form an 'epistemically perfect' whole.

16.2.9 Closed Quantum Systems Involve an Infinity, which Cannot be Supported

If the description of closed quantum systems necessarily involves an infinity (page 299), and infinity cannot be supported in the physical world (page 300), it follows that closed quantum systems cannot exist in the physical world. This is the third reason why they cannot exist (cf. page 295). As far as I see, the argument mainly concerns the relation of this world with certain abstractions, namely those involving infinities.

Now, some have argued that infinity could be detectable in the physical world by means of the recent result of Ji et al. (2020). They provide a non-local game whose winning value is 1 if the strategy of the two players is of a certain kind, and at most 1/2 if it is of another kind. The two players are in possession of a quantum system each, such that the strategy of the first kind can be implemented

only if their quantum systems are of infinite dimension. Hence, if it were possible to measure the value of this game, and the result were such that one could conclude it is 1, it would follow that their respective dimension was infinite. This would amount to witnessing infinity in the physical world! There are many challenges (some perhaps unsurmountable) to realise and measure such a physical system. But, if it were possible to do so, my third reason would be wrong. And, among others, the Church–Turing thesis would become obsolete.

Let us now turn to our house, the universe.

16.3 The Universe as a Plurality

Can you imagine reaching the end of the universe and stretching an arm outwards? And travelling back in time to the very first moment of the universe, and then stepping further back to the moment before? Can you imagine conceiving the inconceivable? And thinking the unthinkable?

All these contradictions have a common skeleton: they try to reach beyond a limit from within that very place. On the one hand, they are within a place—be it physical or non-physical, like space, time, the set of caused events, or of thinkable things. On the other hand, they aim to reach beyond this place, to exit it, to transcend it—to stretch an arm beyond the edge, to exist before the beginning of time, to apprehend the inconceivable, to think the unthinkable. The first condition is called *closure* and the second *transcendence* (Priest 2001). To transcend a condition while being subject to it is a contradiction.

Many philosophical questions can be seen through the lens of the tension between transcendence and closure (Priest 2001). Many thinkers fall into the trap of affirming that something is unreachable, inconceivable, unthinkable, ineffable—and the trap becomes a whirlpool that pulls the logical ground from beneath their feet. For instance, for Parmenides, what is not can neither be nor be thought of—yet here we are, thinking about it at this very moment. If scepticism contends that 'Nothing can be known', then the assertion made by this statement should itself be unknowable. If relativism posits that 'Everything is relative', then this assertion should also be subject to relativity. A statement like 'Something is indescribable' is problematic, because it ascribes a quality ('indescribable') to that entity, thereby contradicting the claim that it cannot be described at all. For example, Kant asserts that the noumenon cannot be predicated, but we just predicated it—asserting that there are entities beyond our judgement is, in itself, a form of judgement about them. Similarly, Aristotle's prime matter is all in potentiality and nothing in particular, it is pure possibility, contains all properties while itself having no property—which is contradictory because we are currently assigning a property to it. Or Quine argues and communicates that meaning is inscrutable,

yet this statement should make this very communication impossible. The very adjective 'ineffable' embodies a contradiction.

This taste of contradictions illustrates the many manifestations of transcendence and closure. Priest identifies them at the limits of cognition, conception, expression, and iteration. My aim was not to quote these thinkers with precision, but rather to demonstrate the many ways how one might attempt to reach the end of some universe and stretch out the arm—this absurdity proposed by Archytas has many reincarnations.

I will rely on the failure to conceive of an ultimate closure in order to argue that the universe should be conceived, if anything, as a plurality, and in particular, not as a closed quantum system. Let us start by examining the skeleton of the tension.

16.3.1 Transcendence, Closure, and Existence

The tension is structured as follows:

Something is contained within a set	(Closure)
This something transcends this set	(Transcendence)
This something exists	(Existence)

As a set, we can imagine the collection of things with a certain property. This skeleton is called the *Inclosure Schema* or *Schema T* (Priest 2001, Bolander 2017).

Transcendence, Closure, and Existence imply a contradiction. In mathematics or any study built on logic, contradictions tend not to be admitted. This is mainly due to the principle of explosion in consistent logics, by which a contradiction allows to derive any statement as true: if *A* and not *A*, then everything follows. Therefore, if we find a contradiction, we reject it. To do so, we conclude that at least one of the three (C, T, or E) is false—usually Existence, as the others tend to be true by construction. Some, notably Priest, appeal to accepting the contradiction and adopting a so-called paraconsistent logic, where the principle of explosion does not hold. But the prevailing approach is to deny the existence of the object that makes the contradiction possible. We will see examples in matters of infinity below.

Yet, the mandates of logic may not be as rigid in philosophy, or may be open to scrutiny. How we traverse the muddy terrain without disappearing in the mire will depend on the specific question, but at the core we will have the problem that to be conceived and to be bounded are essentially the same thing (Priest 2001). Because if something is bounded then it can be conceived as everything that is on this side of the boundary, and if something is conceived then it is bounded by the terms of the conception. In a very naive example, if the Earth were flat, envisioning the magnificent edge that bounds the oceans would involve, even if

just implicitly, contemplating what may lie beyond this boundary. I think that art, especially, confronts contradiction by forsaking the claim to universality.

Let us now see how to stretch out an arm at the end of the universe.

16.3.2 I am a Liar

A systematic way of escaping closure and transcending is through self-reference and negation. If I say:

I am a liar

you will conclude that I am lying if and only if I am not lying. Similarly, if we consider the statement:

This sentence is false

we will conclude that it is true if and only if it is false. If we analyse this sentence with the same tools used to formulate it, we are not able to give it a consistent value of truth or falsehood. In this sense, this sentence transcends the set of true or false sentences. Similarly, Gödel demonstrated the first incompleteness theorem by constructing a sentence in an axiomatic system that says:

I am not provable. (☽)

We again conclude that we cannot prove or disprove it—it transcends provability. By Gödel's theorem, one of the most profound results in mathematics, formal systems with a minimum level of expressiveness cannot be consistent and complete simultaneously. 'Consistent' means that we cannot prove a false sentence as true, and 'complete' that we can prove all true sentences. Some have concluded that truth is a notion greater than provability; others that our brain is more powerful than a machine because we can grasp the paradox (but this conclusion may be premature; see Fraser et al. 2021)).

More formally, (☽) reads 'The sentence whose number corresponds to this very sentence is not provable'. A crucial ingredient is the assignation of 'handles' or 'names' (called Gödel numbers) to sentences, which then serve as variables, so that sentences can *talk about* sentences. Gödel numbering allows for reference, similarly to how the code of the Turing machine can be written in a tape and fed to another Turing machine, or how sets are given a 'name' or 'tag' in Cantor's diagonalisation argument.[20] If reference is the first step, self-reference is the next one: if reference is expressive enough, it can include itself. Self-reference composed with a function without a fixed point, such as negation, leads to contradictions which tend to result in limitations.

The versatility and scope of self-reference and negation are remarkable, partly because the paradox cannot be 'fixed' (see, e.g., Bolander 2017). I summarise it with the maxim:

No system can consistently describe all of its own attributes. (☼)

Let me explain (☼) in the simplest case, that of a set, where it coincides with Cantor's theorem. For a system, we take a set S. If we identify an attribute with its extension, i.e., the set of elements where it is true, then an attribute is an element of the set of subsets of S, written 2^S. For example, if S is the set of natural numbers and f is the attribute 'even' so that $f(n) = 1$ if n is even and $f(n) = 0$ if not, we identify f with the set of even numbers,

$$\{2, 4, 6, 8, 10, 12, \ldots\}.$$

Cantor's theorem states that there is no way to pair the elements of S with those of 2^S without leaving an element of 2^S unmatched. S cannot encompass its attributes 2^S, that is, S cannot thoroughly and consistently talk about itself. *No* S can encompass 2^S—no matter how wild and intricate S is, 2^S will be more so. It follows that there are infinitely many types of infinities! If one infinity blows our mind as if you were touching the stars, imagining infinitely many types of infinity is like flying to the end of the universe and coming back to explain it.

So Cantor's theorem transforms the ancient liar paradox into a technique called diagonalisation, which, for sets, allows tearing any membrane of purported closure and transcending it. It implies that there is no such thing as the set of all sets, there is no largest infinity, there is no ordinal that contains all ordinals, there is no ultimate closure—one can always transcend (Priest 2001, Moore 2019). For other mathematical objects, Lawvere's fixed-point theorem (Lawvere 1969, Yanofsky 2003, Gonda et al. 2024) is, as far as I know, the most precise and powerful formulation of the scope of transcendence and closure, as it expresses the conditions that a system must satisfy to fall prey to these contradictions. This theorem specialises in Cantor's theorem, Gödel's theorem, the uncomputability of the halting problem, Russell's paradox, and more, which express deep and irreparable limitations in provability, computability, the notion of set, learning, etc. All have at their core a contradiction of self-reference and negation that forces them to conclude that certain objects cannot exist. The consequences are devastating across mathematics, computer science, and any enterprise nourished by logic.

What are the consequences for the physical universe? Does the physical world abhor contradiction? That depends on the relation between the formal and the physical. A poignant example is provided by the grandfather's paradox, which logically amounts to self-reference and negation in time. Actions such as killing your

past self are prohibited by causality. Yet, what comes first? Causality or the avoidance of paradoxes? What is more primitive, the flow of time leading to causality or nature's 'need' to avoid paradoxes? If logics should reflect some relations of the physical world, then causality; if it is nature that instantiates mathematical relations (held together by a consistent logic), then the avoidance of paradoxes ought to be more primitive.[21]

Be that as it may, a pragmatic standpoint will support the view that the avoidance of paradoxes is important for the physical world too. The aforementioned enterprises nourished by logic have not only led to a very deep understanding of our world but also triggered the development of many disruptive technologies. The point is that they are useful to model the world and to devise ideas to manipulate it. If such enterprises avoid paradoxes, by this a posteriori argument, it seems that the physical universe does too.

Now forget everything I just said and consider (☼) again. It seems to me that I can think about myself without encountering insurmountable walls, and don't you think you can do the same? By 'think' I mean reflect—I can engage in self-reflection and share these very thoughts with you. That does not mean that I can consistently and completely describe myself—in particular, I cannot achieve 'self-transparency', partly because reflection trails back on experience.[22] So the human condition seems to relate to the limitation of (☼) in a surprising way. I have argued elsewhere (De les Coves 2024) that we carry the contradiction between transcendence and closure in the syntax of our being, that being human consists of struggling to enclose infinity in a finite body, that we transcend and enclose ourselves—that the person, the *prosopon*, the one behind the mask, is both part and whole at the same time.

Let us now tear down the hierarchies and entangle them.

16.3.3 Tangled Hierarchies

Graphically, self-reference and negation can be understood as a tangled hierarchy, which is a hierarchy that 'closes in on itself', meaning that the highest level ends up being influenced by the lowest level. In fact, the demonstration of Lawvere's theorem—which is very short—consists of a diagram that captures the tangled hierarchy generating the contradiction. Hofstadter (1979) identifies in *Gödel, Escher, Bach* this tangled hierarchy in Gödel's proof, works of Escher, and some works of Bach, where the 'ornaments' of the piece of music evolve into the primary structure. In formal systems, the axioms and transformation rules are expected to determine the truth or falsehood of any well-formed sentence in that system. Yet, Gödel's sentence (☼) says something meaningful about the very axioms that give rise to it (in particular, it expresses a limitation). Something remarkable has occurred: reference has emerged within the axiomatic system.

Gödel's is not a sterile sentence whose truth can be proven or disproven from the axioms, but it *points* somewhere, in particular it refers to itself.[23]

Surprisingly, some aspects of the functioning of DNA form a tangled hierarchy, too (Hofstadter 1979). Because DNA is expressed in proteins and determines their composition, it seems to be at a higher hierarchical level. Yet some proteins determine which segments of the DNA are expressed—so they seem to be at a higher level. We find a tangled hierarchy in the reading and translation of the alphabet that materially conforms us, in the smallest units (of meaning) of our very packaging.

In formal sciences, entangling a hierarchy is 'bad' because formal systems form a hierarchy where everything is derived from the (unjustified) truth of axioms. (Imagine a world resting on a giant turtle that is supported by nothing—this is the role of axioms). If we entangle the hierarchy and find a contradiction, we conclude that some assumption is false, for example, the existence of an object. But in other areas, entangling a hierarchy can be good or magical, because it may reveal a mechanism of emergence, as if rising from the ground by pulling one's own shoelaces. In fact, the tangled hierarchy is Hofstadter's proposal to explain the emergence of the self and consciousness, as explained more clearly in *I Am a Strange Loop* (Hofstadter 2007).

The tension between transcendence and closure should be expressible as a tangled hierarchy, where transcendence corresponds to the higher level and closure to a lower one. This connection would underline the idea that we must embrace contradictions, since, if Hofstadter is right, they are very close to the self—so close that they logically constitute us.

16.3.4 Infinity: The Part Embraces the Whole

What defines an object? What parts make up a whole? These very significant questions (in my opinion) occupy metaphysics (see, e.g., Carroll & Markosian 2010).

Consider a whole and a proper part of this whole, that is, a part that is neither empty nor the whole itself—for example, a part of a leg, a part of a fish, or a part of a rainbow. By definition, a proper part is a subset of the whole, in the sense that there are elements of the whole that do not belong to the proper part.[24] It seems evident that the elements of the proper part cannot be paired with those of the whole, because the proper part is *smaller* than the whole, so the part cannot encompass the whole.

In the infinite, a proper part can encompass the whole. Formally, each element of the part can be paired with an element of the whole. For example, in the natural numbers, we can pair each even number $2n$ with a natural number n and leave no natural number unpaired, because *there is no last element*. This is very bizarre: a part of a mountain is different from the whole mountain, and the same goes for a

cow, a nail, or any physical thing. In the infinite, a proper part can be equivalent to the whole (in a formal sense). The whole transcends the part in the sense that there are elements of the whole that are not in the part (such as the odd numbers), but at the same time, the part achieves closure because it encompasses the whole (with the pairing). In fact, the existence of a surjection from a proper part to the whole is a hallmark of infinity. Because infinity has 'no boundary', such outlandish things can happen. Priest calls it a paradox at the limit of the iterable.

16.3.5 The Universe as a Plurality

The most profound lesson of transcendence and closure is that there is no ultimate closure, or, if there is one, we cannot conceive of it because we would have to imagine a membrane around it, which would brush against the very thing we are denying. Lawvere's theorem provides a systematic way to open a door of the purported final house and step out into the street. The collection of houses (closures) and streets (their transcendence) cannot be seen as a new house; there is no such thing as the set of all sets. *There are multiplicities that cannot be conceived as a unit*—they cannot be understood as a whole and therefore cannot be considered 'a thing'. Cantor called these multiplicities inconsistent totalities; I call them streets and houses that cannot be grouped in a new house.

Some say that the only closed system that truly exists is the universe as a whole. I feel very uncomfortable making any such claims. If anything, I want to push against the hopes (or aspirations) that reality should be conceived of as a unity and not a plurality. If a set is a many which allows itself to be thought of as one, and some collections are too large to be considered as one, how could the universe be considered as one? We should not hope for a last refuge, an ultimately closed system. On the epistemological side (and almost ethical side too), this stance is grounded in my finiteness, which instils in me a profound sense of humility. It makes me sympathetic to the view that, if the universe can be modelled as anything (an extremely ambitious conditional), it should be according to the open systems view (Cuffaro & Hartmann 2021). But the question whether the universe should be described as a closed or open system is probably moot, as both views presuppose that quantum mechanics holds at the scale of the universe—an unlikely fact, in my eyes.

I disagree with Einstein (2023) when he says:

> We must escape the chaos of contradictory fragmentary events and futile passions that produce nothing but unrest. We can only free ourselves from the chaos around us by creating rational scientific or ethical systems … Only with reason can we free ourselves from transience and, through objectivity, approach what is eternal.

And I agree with Pessoa (2020) (or his heteronym Alberto Caeiro):[25]

> One excessively clear day,
> a day that made one wish to have worked hard
> so as not to have to work at all that day,
> I caught a glimpse, like a path between the trees,
> of what might very well be the Great Secret,
> the Great Mystery that the false poets talk about.
> I saw that there was no Nature,
> that Nature does not exist,
> that there are valleys, mountains, plains,
> that there are trees, flowers, grasses,
> that there are rivers and stones,
> but that there is no whole to which all this belongs,
> that the real and true ensemble
> is a mania of our ideas.
> Nature is made of parts without a whole.
> Perhaps this is the mystery they speak of.
> And it was that, without thinking or pondering,
> I realised must be the truth
> that everyone tries to find and does not find,
> and that only I, by not having sought it, have found.

The universe is too wild to be considered one. We ought to approach such magnificent pluralities with awe, humility, and fantasy.

16.4 Coda

What would physics without the mathematical ideal look like? Del Santo & Gisin (2019) offer some hints. What would physics centred around the real look like? The theory of open quantum systems provides some clues. What would mathematics without infinity be like? It may resemble constructivism. Interestingly, the pivot on the ideal seems to be alien to biology—limits to infinity or asymptotic reasoning do not seem to be of help.

Abandoning the reference to the ideal is, in my eyes, analogous to Popper's shift in epistemology, where the old question 'What is truth?' is replaced by 'How can sources of error be recognised and corrected?' Instead of referring to the ideal, he places himself 'in the middle of the sequence' and enquires how to improve his knowledge. This gives rise to falsifiability, which, ironically, is our methodological paradigm. Similarly, 'What is justice?' is replaced by 'How can we improve the political system without violence?' This gives rise to democracy, presumably.

Eschewing the appeal to the ideal, in matters of justice, is a move whose importance cannot be overstated. History has told us that justice ideals (utopias) can be highly non-unique and very different for different people.

What is a mystery to me is how we—confused, perishable, open systems—can conceive of the ideal, for example, of infinity. But this mystery may again be just be a different guise of the power of the mind to abstract and to unify, exercised whenever reference takes place.

What is clear is that there is no knowledge without a conceptual framework. Which is why initiatives like this book are so important, because we should not take anything for granted, let alone what is supposed to be fundamental.

Acknowledgements

I wish to thank two anonymous reviewers for their very constructive comments, which have helped me improve this chapter.

Endnotes

1. Interestingly, Tegmark (2008) starts from essentially the same assumption, in his case the existence of a physical reality completely independent from us humans, and draws a radically different conclusion from this chapter, namely that the external physical reality is a mathematical structure. In order to ban mathematical contradictions from existence, he introduces the Computable Universe Hypothesis. More generally, while mathematical monists will be unimpressed by the arguments in this chapter, they also need to face the mathematical contradiction, and precisely because they only posit the existence of mathematical structures, their view is completely tied to the limitations of mathematics and computation.
2. Although I will later adopt an intertwined view of ontology and epistemology, see fn. 13.
3. In Barad's (2007) onto-epistemology, knowledge and knowledge acquisition are given a very different account than this one, yet the first statement of this paragraph would hold in that view too.
4. I consider finitely many dimensions; similar considerations apply to infinitely many dimensions.
5. For example, the qubit channel $\varepsilon(\rho) = (\mathrm{tr}(\rho)I + \rho^t)/3$ cannot be divided even once into two equal channels. This follows from the fact that its determinant is minimal (Wolf & Cirac 2008).
6. In fact, a size bounded by the dimension of the original space suffices. The significance of this fact will be discussed in the light of mereological relations (page 275).
7. Although such ghostly fulcrums are found elsewhere in physics, e.g., in thermodynamics.
8. An extreme form of reductionism can be found in mathematical monist accounts such as Tegmark (2008). Interestingly, Tegmark ends up questioning the use of real numbers to represent physical quantities precisely because of computability issues. Ironically, both Tegmark and this chapter start from a common ground (assuming the existence of an external physical reality), each derives very different conclusions, both wrestle with paradoxes, and end up denying the validity of real numbers to describe a physical quantity (my words) or the existence of a parameter described by a real number (presumably Tegmark's words).
9. Although these separations seem to disappear in the approximate case (De las Cuevas et al. 2021).
10. Recently generalised in De les Coves et al. (2023).

11. The potential infinity can be imagined like a clock endlessly ticking. Its ticking is potentially but never actually infinite. The ticking is in a constant state of becoming, but never actually *is*, in its entirety. It never achieves full being (Moore 2019).
12. To dispel suspicions that the argument may be circular, consider the statement that the universe is finitely old (13.7 billion years old). The age of the universe can be determined by measuring the Hubble constant today and extrapolating back in time. In such calculations variables such as time and distance are represented by real numbers; more precisely, an amount of time or space is represented by a real number times seconds or metres, respectively. Both the definition of a second and a metre are based on observable events, so they do not involve infinites. (One second is the duration of 9192631770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the Caesium 133 atom, and one metre the length of the path travelled by light in vacuum during a time interval of 1/299792458 of a second.) Thus, to represent time as a real number times seconds is an unnecessary concession to infinity—it suffices to represent it as rational with finite precision times seconds. That this precision improves with our understanding is perfectly fine, as it does not become arbitrary; the same can be said about distance. Finally, manipulations of such time and distance amounts involve taking fractions, so do not involve an infinity. A similar argument can be made about the statement that the universe is finite in space. This point is also made on page 301. What we need, I believe, is a notion of 'infinite for all practical purposes', see endnote 17.
13. Implicitly, I am adopting a view where the ontology depends on the epistemology, as I do not want to grant that anything exists if I (or anyone else) cannot come to know it. This view could find a philosophical underpinning in Barad's (2007) onto-epistemology of techno-scientific practices.
14. Please see Moore (2019: Ch. 1) for the beautiful and moving ideas on infinity by Anaximander, Pythagoras, Parmenides, and Plato.
15. Why should the fantastic structure of complex numbers help us describe the quantum world? 'Fantastic' because it was conceived as something imaginary which required the introduction of a symbol *i* to solve certain algebraic equations. Why the minuscule world should be so beautifully described by this mathematical construct is a mystery to me.
16. Renou et al. (2021) take it for granted that a quantum system should be described by an element of a real or a complex Hilbert space, both of which involve an uncountable infinity. Being a Hilbert space guarantees that sequences converge, which is a mathematical requirement that is (I argue) generally physically inaccessible. For example, I disagree with their statement that 'physics experiments are described by the statistics they generate, that is, by probabilities, and hence real numbers'. We can only access relative frequencies, which are rational numbers of finite precision. We need to grant the existence of the potential infinite to repeat an identical experiment an unbounded number of times in order to provide a relative frequency with arbitrary precision. I do not see how an irrational number (algebraic or transcendental) can be accessed physically. And one would still need to guarantee that such experiments are identical.
17. What is ultimately at stake here is the fact, mathematically, that it is meaningful to distinguish finite from unbounded, and unbounded from infinite, but these distinctions do not seem to be physically meaningful. Consider again the statement 'The universe is 13.7 billions years old'. As far as the mathematical distinction is concerned, the finite number quantifying the amount of time is equivalent to any other finite number, like five. Yet, physically, this number is not only large but there is a sense in which it is the largest. This sense is completely missing in the mathematical theory. Physics would need a notion of 'infinite for all practical purposes', which may also allow to formalise independence in emergence.
18. < and > denote, as usual, 'is smaller than' and 'is larger than', respectively.
19. The inadequacy of reductionism is one of the driving forces behind the research field called Complex Systems. See, e.g., Solé & Goodwin (2000), Solé (2009), Holland (2014). One of the most striking examples of the failure of reductionism is the brain itself: we understand well how pairs of neurons behave, but how does cognition or consciousness emerge?
20. This pattern can be made explicit by means of the simulator of Gonda et al. (2024). The emergence of reference amounts to a transfer of information from the target T to the context C; this is

radical for Turing machines, where a single object *T* (the universal Turing machine) can run any algorithm.

21. Since Tegmark (2008) identifies physical reality with mathematical reality, he would presumably consider the avoidance of paradoxes the only criterion. However, his account of causality is unclear to me.
22. This temporal structure of mental acts is crucial to argue how our brains makes sense of the liar paradox in Fraser et al. (2021).
23. Hofstadter calls it the emergence of meaning, but I find 'reference' more accurate.
24. While this may appear obvious to us, it may not be so for other thinkers. For example, Leibniz postulated a very awkward mereology of the world where everything is composed (in a nuanced way) of monads, and each monad mirrors the whole universe (see, e.g., Antognazza 2016).
25. Twenty years ago I read several books of Pessoa's heteronyms and selected only this poem, I don't know why. Recently, I encountered it again and it acquired an ironically consistent meaning within my understanding of transcendence and closure.

Bibliography

Antognazza, M. R. (2016), *Leibniz: A Very Short Introduction* (Oxford University Press).

Barad, K. (2007), *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning* (Duke University Press).

Bolander, T. (2017), 'Self-Reference', in E. N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University). Fall 2017 edition. URL https://plato.stanford.edu/archives/fall2017/entries/self-reference/.

Carroll, J. W., & Markosian, N. (2010), *An Introduction to Metaphysics* (Cambridge Introductions to Philosophy).

Cubitt, T. S., Perez-Garcia, D., & Wolf, M. M. (2015), 'Undecidability of the Spectral Gap', in *Nature* 528: 207.

Cuffaro, M. E., & Hartmann, S. (2021), 'The Open Systems View', arXiv:2112.11095.

De las Cuevas, G., Drescher, T., & Netzer, T. (2020), 'Quantum Magic Squares: Dilations and their Limitations', in *Journal of Mathematical Physics* 61: 111704.

De las Cuevas, G., Klingler, A., & Netzer, T. (2021), 'Approximate Tensor Decompositions: Disappearance of Many Separations', in *Journal of Mathematical Physics* 62: 093502.

De las Cuevas, G., Schuch, N., Pérez-García, D., & Cirac, J. I. (2013), 'Purifications of Multipartite States: Limitations and Constructive Methods', in *New Journal of Physics* 15: 123021.

De les Coves, G. (2024), 'Being Human: The Wound of Infinity. An Intimate Encounter of Science and Culture', in R. Garrigasait & J. Ramoneda (eds.), *In an Open Field: Contemporary Cultural Practices* (Barcelona: Institut Ramon Llull).

De les Coves, G., Corominas-Murtra, B., & Solé, R. (2024), 'Universality and Complexity in Natural Languages: Mechanistic and Emergent'. URL https://doi.org/10.20944/preprints202402.1330.v1

De les Coves, G., van der Eyden, M., & Netzer, T. (2023), 'Beyond Operator Systems'. arXiv:2312.13983.

Del Santo, F., & Gisin, N. (2019), 'Physics Without Determinism: Alternative Interpretations of Classical Physics', in *Physical Review A* 100: 062107.

Eco, U. (1997), *Kant and the Platypus: Essays on Language and Cognition* (Harcourt).

Einstein, A. (2023), *The World As I See It* (Zinc Read).

Fankhauser, J., Gonda, T., & De les Coves, G. (2024), 'Epistemic Horizons from Deterministic Laws: Lessons from a Nomic Toy Theory', in *Synthese* 205: 136.

Fraser, P. T., Solé, R., & De las Cuevas, G. (2021), 'Why Can the Brain (and Not the Computer) Make Sense of the Liar Paradox?', in *Frontiers in Ecology and Evolution* 9: 802300.

Gefter, A. (2023), 'How to Unsplit the World: Quantum Mechanics, Cognitive Science, and the Subject–Object Divide', *Foundational Questions Institute Prize Essay*. URL https://forums.fqxi.org/d/4037.

Goldblatt, R. (1998), *Lectures on the Hyperreals: An Introduction to Nonstandard Analysis* (Springer).

Gonda, T., Reinhart, T., Stengele, S., & De les Coves, G. (2024), 'A Framework for Universality in Physics, Computer Science and Beyond', in *Compositionality* 6: 3.

Hofstadter, D. (1979), *Gödel, Escher, Bach: An Eternal Golden Braid* (Basic Books).

Hofstadter, D. (2007), *I Am a Strange Loop* (Basic Books).

Holland, J. H. (2014), *Complexity: A Very Short Introduction* (Oxford University Press).

Horn, R. A., & Johnson, C. A. (1985), *Matrix Analysis* (Cambridge University Press).

Ji, Z., Natarajan, A., Vidick, T., Wright, J., & Yuen, H. (2020), 'MIP*=RE'. arXiv:2001.04383.

Klingler, A., van der Eyden, M., Stengele, S., Reinhart, T., & De las Cuevas, G. (2023), 'Many Bounded Versions of Undecidable Problems are NP-Hard', in *SciPost Physics* 14: 173.

Lawvere, F. W. (1969), 'Diagonal Arguments and Cartesian Closed Categories', in *Lecture Notes in Mathematics* 92: 134.

Loux, M. J., & Crisp, T. M. (2017), *Metaphysics: A Contemporary Introduction* (Routledge Contemporary Introductions to Philosophy). 4th edition (New York: Routledge).

Lupini, M., Mančinska, L., & Roberson, D. E. (2020), 'Nonlocal Games and Quantum Permutation Groups', in *Journal of Functional Analysis* 279: 108592.

Moore, A. W. (2019), *The Infinite* (Routledge). Third edition.

Pessoa, F. (2020), *The Complete Works of Alberto Caeiro: Bilingual Edition* (New Directions). Trans. M. J. Costa.

Popper, K. (1963), *Conjectures and Refutations: The Growth of Scientific Knowledge* (Routledge).

Priest, G. (2001), *Beyond the Limits of Thought* (Clarendon Press).

Reinhart, T., & De las Cuevas, G. (2025), 'The Grammar of the Ising Model: A New Complexity Hierarchy'. in *Proceedings of the Royal Society A* 481: 20240579.

Renou, M.-O., Trillo, D., Weilenmann, M., Thinh, L. P., Tavakoli, A., Gisin, N., Acin, A., & Navascues, M. (2021), 'Quantum Physics Needs Complex Numbers', in *Nature* 600: 625–9.

Solé, R. (2009), *Redes Complejas: Del Genoma a Internet* (Tusquets).

Solé, R. V., & Goodwin, B. (2000), *Signs of Life: How Complexity Pervades Biology* (Basic Books).

Stengele, S., Drexel, D., & De las Cuevas, G. (2023), 'Classical Spin Hamiltonians are Context-Sensitive Languages', in *Proceedings of the Royal Society A* 479: 20220553.

Tegmark, M. (2008), 'The Mathematical Universe', in *Foundations of Physics* 38: 101.

Van der Eyden, M., Netzer, T., & De las Cuevas, G. (2022), 'Halos and Undecidability of Tensor Stable Positive Maps', in *Journal of Physics A* 55: 264006.

Wigderson, A. (2019), *Mathematics and Computation: A Theory Revolutionizing Technology and Science* (Princeton University Press).

Wolf, M. M., & Cirac, J. I. (2008), 'Dividing Quantum Channels', in *Communications in Mathematical Physics* 279: 147.

Wolf, M. M., Cubitt, T. S., & Perez-Garcia, D. (2011), 'Are Problems in Quantum Information Theory (Un)decidable?'. arXiv:1111.5425v1.

Yanofsky, N. S. (2003), 'A Universal Approach to Self-Referential Paradoxes, Incompleteness and Fixed Points', in *Bulletin of Symbolic Logic* 9: 362.

17
Perspective Duality as a Physical Requirement

Wayne C. Myrvold

17.1 Introduction

There is a tension lurking in the literature on philosophy of physics, between two strands of thought. One strand, exemplified in J. S. Bell's 'Against Measurement' (Bell 1990), contends that a fundamental theory ought to be capable of being formulated without a split between *system* and *environment*. Any such split, the thought goes, is a somewhat arbitrary distinction imposed by us, for convenience in our modelling, and corresponds to nothing in physical reality—hence, when speaking of the really deep issues, the questions about what the world is like, at the fundamental level, it must be dispensed with. This was, in part, a reaction against the insistence, from those infected with the Copenhagen virus, that quantum mechanics requires for its formulation a split between system and apparatus, with the apparatus treated classically, or worse, a split between system and *observer*. Against this, it has been emphasised that there are formulations of quantum theory, such as dynamical collapse theories and the de Broglie–Bohm pilot-wave theory, that require no such split, and can be taken to apply to both macroscopic and microscopic objects, to both system experimented on and experimental apparatus, modelling the interactions between the two *within* the theory.

The other strand of thought is also found in Bell.

> A respectable class of theories, including contemporary quantum theory as it is practiced, have 'free' 'external' variables in addition to those internal to and conditioned by the theory. These variables are typically external fields and sources. They are invoked to represent experimental conditions. (Bell 1977: 81; 2004: 101)

To treat a variable as *external*, or *exogenous*, is to treat its value as simply given, not to be subjected to the dynamical laws of the theory being used to model the dynamics of the system under investigation. Whether or not a given parameter is exogenous is thus not a property of the physical situation, but of how it is

Wayne C. Myrvold, *Perspective Duality as a Physical Requirement*. In: *Open Systems: Physics, Metaphysics, and Methodology*. Edited by: Michael E. Cuffaro and Stephan Hartmann, Oxford University Press.
© Oxford University Press (2026). DOI: 10.1093/9780198929277.003.0018

being treated in a given investigation. One and the same quantity may be treated exogenously in one investigation and endogenously in another. As an illustrative example, consider the familiar progression of textbook treatments of planetary motion. As a first pass, one considers the motion of a planet subjected to the sun's gravitation, taken as a fixed gravitational potential unaffected by the planet. On such a treatment, the sun's position is not treated as a dynamical variable, subject to the laws of motion, but is treated as fixed. As a next step one includes the sun's position among the dynamical variables considered, and deals with the system as a two-body problem.

Exogenous variables—or, we should say, variables treated exogenously, to emphasise that the distinction is in our treatment, not in the nature of things—play two crucial, and related, roles in science. In statistical hypothesis testing, the choice of experiment, or choice of intervention on the system, is treated as an exogenous, or, as Bell would say, a free variable, and this is absolutely essential to statistical inference. In the causal modelling literature, it is common to take an interventionist approach, on which the effects of a change of a variable are studied by regarding the value of that variable as set exogenously.

This poses a problem. If the theory under investigation is a comprehensive theory of fundamental physics, applicable in principle to everything, what sense can we make of treating certain variables as exogenous? If the notion of *intervention* is essential for the application of causal concepts, does this mean that, as Pearl (2000: 350) put it, when the subject matter is the whole of reality, 'causality disappears because interventions disappear—the manipulator and the manipulated lose their distinction'? And, if our experimental apparatus, including any devices (such as randomisers) used to determine the choice of treatment, are subject to the dynamical laws of our theory, what sense does it make to treat the outputs of such devices as 'free' variables? Some will say that this, too, makes no sense when fundamental physics is under investigation.

Things are not so dire, I will argue. The apparent tensions can be resolved, and, in their resolution, we gain insight into physical theorising. In what follows, I will argue for what I call *perspective duality*. Bell is right that any candidate for a fundamental physical theory worth its salt should admit a formulation without a system–environment split. That is, it should be possible to adopt a *totalising* perspective, an ideal point of view that would leave nothing out. But also, to avoid the threat of making nonsense of statistical hypothesis testing, and of causal reasoning, a physical theory worth its salt should *also* admit of a formulation with exogenous variables. We should be able to adopt a non-totalising, *limited* perspective. As we will see in section 17.7, the possibility of switching between these two perspectives, consistently, places a constraint on what a reasonable physical theory can be like.

17.2 Comprehensive Theories with Deterministic, Invertible Dynamics

The theories of the sort we'll be interested in are those with certain features that are usually assumed by those who imagine that they know what a fundamental physical theory would look like, if we had one.

The first is that there be no built-in limitation in the scope of a theory. Unlike theories of the special sciences, which deal with limited domains of reality, the theories we will be concerned with will be capable, at least in principle, of treating of the behaviour of everything that there is. They will also (unlike, say, a theory of gravity or of electromagnetism) deal with all sorts of interactions. Call this feature *comprehensiveness*.

The dynamical laws of the theory will be assumed to be *deterministic*, and *invertible*. *Determinism* is the condition that a complete specification of the state of the world up to a given time, together with the theory's dynamical laws, uniquely determines its state at all future times. That is, if two dynamically possible complete histories of the world agree at given time t, they agree at all times to the future of t. Another way of saying this is that, if two complete histories of the world differ at some time t, they differ at all earlier times. In a relativistic spacetime, talk of the state at a time is to be replaced by talk of the state on a spacelike slice of spacetime that is a *Cauchy surface*, meaning that every past- and future-inextendible timelike curve intersects it exactly once. (We assume that we are dealing with a globally hyperbolic spacetime, that is, one that admits of Cauchy surfaces; otherwise, talk of determinism fails to make sense.)

Invertibility is the temporal inverse of determinism; it is the condition that a complete specification of the state of the world at a time, together with the theory's dynamical laws, uniquely determines its state at all past times. That is, if two dynamically possible complete histories differ at some time t, they differ at all times to the future of t; there is no forgetting of the past.

Determinism is often taken as the conjunction of what we are calling 'determinism' and 'invertibility'. For our purposes it is useful to distinguish them, as the two conditions play distinct roles.

17.3 Physics and Causal Structure

In some quarters, at least, the causal structure of a physical theory is regarded as an important feature of the theory. In particular, having a causal structure compatible with relativistic spacetime structure is often taken to be a *desideratum* for a proposed physical theory, perhaps even a necessary condition for the theory to be taken into serious consideration. Relativistic structure (whether special

or general) has the following features. The past of any event e is the set of events on or within its past light-cone, that is, the set of events from which a signal or other influence propagating at the speed of light or slower can reach the event. The future of event e is the set of events on or within its future light-cone, that is, the set of events that can be reached from e by a signal propagating at light-speed or slower. All other events are *spacelike separated* from e, neither to the past of e nor to the future of it. If, as is usual, we take the cause–effect relation to be a temporally asymmetric one, with causes temporally preceding their effects, compatibility with relativistic spacetime structure requires there to be no cause–effect relations between spacelike separated events. (Absent this assumption about temporal asymmetry, there is no reason to hold that relativity prohibits superluminal signalling or other cause–effect relations between spacelike separated events.) The constraint of relativistic causality is distinct from, and logically independent of, the requirement that the theory's dynamical laws be invariant under the symmetries of Minkowski spacetime, a requirement often called *Lorentz covariance*.

In the case of a classical field theory, the requirement of relativistic causality is taken to amount to the requirement that the values of a field in a given spacetime region be determined by the values it takes in its past light-cone, and that field values in two spacelike separated regions can be specified independently of each other. In the case of a quantum field theory, the requirement is that the theory satisfy a condition known as *microcausality*.

The physical content of the microcausality condition is the demand that a choice of what, if any, experiment to perform not affect the probability of outcomes of an experiment at spacelike separation. It is cashed out mathematically as follows.

As in nonrelativistic quantum mechanics, in a quantum field theory one associates operators with the dynamical variables ('observables') of a system. These operators can be added to each other, and one can multiply operators by scalars and by each other. Subject to appropriate conditions on these mathematical operations, the set of operators associated with a physical system form an *algebra*. In quantum theories the multiplication operation is, in general, noncommutative, meaning that the product of two operators may differ depending on the order. That is, for two operators $\hat{A}$, $\hat{B}$, $\hat{A}\hat{B}$ may differ from $\hat{B}\hat{A}$. This is what distinguishes quantum theories from classical ones. Of course, for some operators, the product could be independent of order. When $\hat{A}\hat{B}$ is equal to $\hat{B}\hat{A}$, the operators $\hat{A}$ and $\hat{B}$ are said to *commute*.

In a relativistic quantum field theory, for any region of spacetime with nonempty interior, there is a set of observables corresponding to outcomes of experiments that can be performed in the causal closure of that region. We will call the algebra of operators containing the operators corresponding to observables associated with a spacetime region $\mathcal{O}$, $\mathfrak{A}(\mathcal{O})$.

One typically models experiments as follows. It is assumed that, associated with the experiment to be performed, there is a set of operators $\{\hat{M}_i\}$, satisfying the

condition that $\sum_i \hat{M}_i^\dagger \hat{M}_i$ is the identity operator, such that, when the experiment is performed on a quantum state represented by a vector $|\Psi\rangle$, the result is to produce, for some i, the state $\hat{M}_i|\Psi\rangle$, with probabilities proportional to $\|\hat{M}_i|\Psi\rangle\|^2$. This is the condition that the state-transformation correspond to what is called a *nonselective completely positive operation*.

The microcausality condition requires that these goings-on be irrelevant to probabilities of an experiment regarding some observable B, performed at space-like separation from the A-experiment. And that means that, when computing probabilities for outcomes of a B-experiment, it shouldn't matter whether we use the original state vector $|\Psi\rangle$, or a weighted average of the probabilities computed using state vectors $\{\hat{M}_i|\Psi\rangle\}$, weighted by the probabilities of the corresponding outcomes. It is not difficult to show that the operation always leaves the probabilities of a measurement of an observable B unchanged if and only if the operator $\hat{B}$ commutes with every one of the operators $\{\hat{M}_i\}$ (see the appendix of Myrvold 2021 for a simple proof).

The constraint of relativistic causality also has implications for proposed modifications of the usual quantum dynamics. One approach to the so-called 'measurement problem' of quantum mechanics involves a modification of the dynamics to suppress superpositions of macroscopically distinct states. It can be shown that the no-signalling condition places non-trivial constraints on such modifications (see Gisin 1989, Simon et al. 2001).

17.4 The Problem

Randomisation posits employed in statistical hypothesis testing involve a division between the system under investigation and the randomising device used. In the literature on causal inference, the notion of *intervention* plays a role, which involves treating the value of the variable intervened upon as exogenously given.

Some theories are inherently of limited scope, and may be such that we would not and could not expect them to have anything to say about the behaviour of the randomising device and other experimental apparatus. A comprehensive physical theory has no such excuse.

Moreover, for a system that is not acted upon from the outside, a solution of the equations of motion specifies what happens at *all* times, past, present, and future. It is not clear what it would mean for one part of the system to *change* the state of another, or to talk of cause–effect relations, or sender–receiver signalling relations.

Take, for example, the case of signalling. We usually think of Bob, the sender, as having a choice of signals to send, and of Alice, the receiver, as learning, upon receipt of the signal, about the choice that Bob has made. But if Bob and Alice and their apparatus are included in our theoretical description, and if the dynamics are

invertible, then any two solutions in which Bob sends different signals are different at all times to the past of the sending event, and this will include times to the past of Alice's reception. Because of determinism, a state to the past of Alice's reception determines what Alice will receive. That is, a state to the past of both events determines both Bob's choice of signal and what Alice receives. How, then, are we to make sense of talk of Bob's choice of signal having an effect on what Alice receives?

This worry has, in fact, been expressed, in connection with the usual sorts of rationale for the microcausality constraint on quantum field theories. Laura Ruetsche, after introducing a proposed motivation for the constraint, along the lines of what we discussed in section 17.3, writes,

> I take this motivation to fail because I take one of its presuppositions to be incompatible with *other* axioms of Axiomatic QFT. The presupposition is that the way to model an operation performed in the spacetime region $\mathcal{O}$ is by the change of state wrought by acting on the global state $|\Psi\rangle$ with some element A of $\mathfrak{A}(\mathcal{O})$. My tendentious contention is that, whatever one thinks of measurement in nonrelativistic QM, there's no way to square this model with the idea … that the quasilocal algebra $\mathfrak{A}(\mathcal{M})$ admits a state.[1] If $|\Psi\rangle$ is that global state, it's a state for all space and all time, and the notion of changing it by an operation executed within spacetime is nonsense. If $|\Psi\rangle$ is not that global state, we need an account of what state $|\Psi\rangle$ is, an account that lends coherence to the idea that acting on $|\Psi\rangle$ with an element of the local algebra $\mathfrak{A}(\mathcal{O})$ corresponds to carrying out a local operation performable in $\mathcal{O}$. (Ruetsche 2011: 110; see also Ruetsche 2021: 319)

Ruetsche goes on to express further worries about arguments along the lines we have considered, having to do with prohibition of action-at-a-distance. Since such arguments involve modelling a change of the quantum state brought about by an operation local to a region $\mathcal{O}$ via operation on the state vector with an operator in the local algebra $\mathfrak{A}(\mathcal{O})$, the worry expressed in the quoted passage applies also to them.

Similar concerns have been expressed by others, about how to make sense of intervention when the system being modelled is the entire universe; see Pearl (2000: 350), Hitchcock (2007: 53–4), Woodward (2007: 93).

17.5 Wrong Answers

There are a number of ways to respond to this worry. In this section I will present some that strike me as steps in the wrong direction.

The first might be called *exceptionalism*. This is the idea that there are certain kinds of things that are exempt from being treated as physical objects, subject to

physical laws. The most common sort is that conscious beings are endowed with special powers that transcend the laws of physics.

This would be a radical move, unwarranted by anything we know about the behaviour of the bodies (including the brains) of conscious beings, which suggests, on the contrary, that physical principles *can* fruitfully be applied to them. It would also fail to solve the problems raised, as they arise also when considering the effects of one physical system on another, with no conscious beings involved.

Another proposal that might be made is that, though a comprehensive physical theory ought to be capable of embracing *anything*, it makes no sense to talk of simultaneously modelling *everything*; scientific investigation always involves a split between the system under investigation, and the system doing the investigation. This position is less obviously problematic than the position that there are some things that are forever exempt. Nonetheless, as I will argue below, it is more extreme than the logic of the situation dictates, and there is a more modest escape from the difficulties.

Another move is what might be called *eliminativism*. Bertrand Russell famously remarked, more than a century ago, '[t]he law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm' (Russell 1913: 1). In a similar vein, John Norton, more recently, has written:

> Mature sciences, I maintain, are adequate to account for their realms without need of supplement by causal notions and principles. The latter belong to earlier efforts to understand our natural world or to simplified reformulations of our mature theories, intended to trade precision for intelligibility. (Norton 2007: 12)

And Ruetsche:

> Soberly viewed all the special theory of relativity (STR) can be taken to demand of a theory set in Minkowski spacetime is that it exhibit Lorentz covariance. And this demand is met by QFTs in its scope, provided they satisfy the Covariance axiom—*even if they violate the Microcausality axiom!* (Ruetsche 2011: 109)

On this view, causal notions, though they may be useful in everyday life and in the special sciences, are to be dispensed with in fundamental physics.

Is this right? When we were children, we spoke as children, understood as children, thought as children. Now that we have matured, is causality one of those childish things to be put away?

This move strikes me as throwing the baby out with the bathwater. For one thing, as we have seen, causal constraints are fruitfully employed in construction and evaluation of physical theories. The requirement of relativistic causality imposes constraints not entailed by Lorentz covariance, and, in the context of the

closest things we have to fundamental theories, quantum field theories, it has teeth; it requires that the microcausality axiom be satisfied, which, as Ruetsche notes, is a requirement independent of covariance.

It has also been proposed that in the context of fundamental physical theories, independence posits of the sort prevalent in statistical hypothesis testing be abandoned. This proposal has arisen in the context of experimental tests of Bell inequalities, the results of which, on the usual interpretation, are taken to indicate the reality of some kind of nonlocality in the physical world, albeit a sort that is arguably not at odds with the requirements of relativistic causality.[2]

It was noted by Shimony et al. (1976) that there was an assumption required for the derivation of Bell inequalities that up to that point had been implicit but not explicit in Bell's derivations. This is the assumption that the choice of experiment to be performed can be made in such a way that it is statistically independent of the state of the system to be experimented on, an assumption that has been called the assumption of *measurement independence*, or, somewhat misleadingly, the *free will assumption*. In recent literature it is often referred to as the assumption of *statistical independence*, terminology that is less than ideal, as it fails to specify what variables are to be taken as statistically independent. The rejection of the measurement independence assumption has come to be known as *superdeterminism*.

Shimony, Horne, and Clauser (SHC) illustrate the point with a tale about a conspiracy involving manufacturers of experimental apparatuses and experimenters' assistants. The director of the conspiracy concocts a set of experimental data, consisting of a sequence of experimental settings and results obtained. The experimenters' assistants arrange things so that the experiments performed are the pre-arranged ones, and the experimental apparatus is pre-programmed to yield the stipulated results. Obviously, any set of results can be used in such a scheme, and one can obtain violations of the Bell inequalities without making use of any physical nonlocality.

SHC point out that the assumption of measurement independence is not required by any sort of relativistic causality condition.

> After all, the backward light cones of those two acts do eventually overlap, and one can imagine one region which controls the decision of the two experimenters who chose *a* and *b*. We cannot deny such a possibility. But we feel that it is wrong on methodological grounds to worry seriously about it if no specific causal linkage is proposed. In any scientific experiment in which two or more variables are supposed to be randomly selected, one can always conjecture that some factor in the overlap of the backwards light cones has controlled the presumably random choices. But, we maintain, skepticism of this sort will essentially dismiss all results of scientific experimentation. Unless we proceed under the assumption that hidden conspiracies of this sort do not occur, we have abandoned in advance the whole enterprise of discovering the laws of nature by experimentation. (Shimony et al. 1976: 6; Shimony 1993: 168)

The conspiracy need not involve the plotting of agents; one can (perhaps) imagine theories on which all interactions are local but initial conditions are constrained to produce the desired experimental settings and results. In order for a theory of this sort to be empirically equivalent to standard quantum mechanics, this must hold *no matter what* randomisation method is employed. For example, in the experiment of Shalm et al. (2015), the measurement decisions were determined by applying an XOR (exclusive or) operation to three bits from three independent sources, one of which was a pseudorandom sequence constructed from digits of π and binary strings derived from various popular culture sources, including *Monty Python and the Holy Grail*, the *Back to the Future* trilogy, and episodes of *Doctor Who*, *Saved by the Bell*, and *Star Trek*. A hypothetical cause that achieved the observed statistics via correlation between states of the photons studied and the choice of measurements would have to orchestrate the creative processes leading up to the digitised versions of these cultural artefacts in such a way that, when processed in conjunction with the outputs of the random number generators, produced just the right sequence of experimental choices.

SHC are correct that measurement independence assumptions underly virtually all experimental science. Any conclusion whatsoever drawn from statistical hypothesis testing could be undermined by a rejection of measurement independence. For that reason, SHC argue, a methodology that countenances rejection of a measurement independence assumption whenever its employment would lead to an unwanted conclusion amounts to an abandonment of the enterprise of investigating the world by experimentation. Nonetheless, superdeterminism has found some advocates; see Myrvold et al. (2024) for an overview and references.

I think that SHC are right that we should *not* adopt a methodology according to which it is acceptable to conclude a failure of our randomisers to do their job whenever the ordinary canons of statistical hypothesis testing would lead to an unpalatable conclusion, in the absence of independent reason to suspect a failure of randomisation. Such a methodology would be a blunt instrument that could be used to undermine any conclusion one wishes to reject, and would amount to a violation of Peirce's maxim, 'Do not block the way of inquiry' (Peirce 1931: ¶ 135).

17.6 The Solution

To understand how to deal with these worries, let us consider again the concern raised by Ruetsche in connection with quantum field theories. A global state yields expectation values for all observables, distributed over all of spacetime. The idea of changing this state by an operation executed within spacetime, it is said, is nonsense.

The first thing to note is that there is nothing at all distinctively *quantum* about this argument. If cogent, it applies equally well to, say, classical field theories.

A solution of the field equations of a classical field theory is a global solution, yielding field values distributed over all of space and time. We could replace talk of a global state, in quantum field theory, with talk of a global solution of the field equations, in a classical theory, and the worry would retain its full force.

And this gives us a key to the solution of the puzzle. We do, indeed, apply causal notions to classical field theories; part of the content of a course on classical electromagnetism consists of learning how to generate and manipulate electromagnetic waves! This is possible because, though the theory is capable of treating of the behaviour of charges and currents within it, one can also examine solutions of the field equations with exogenously given charges and currents. So, for example, to examine the effect of a given oscillating current as a source of electromagnetic waves, one can compare two solutions: one in which the current is absent, for all time, and one in which it is endowed with a certain history, which might involve being zero up until some time at which it is switched on. One can then consider solutions to the field equations with the current as a source term, selecting retarded solutions, that is, solutions in which the field at a given spacetime point depends on charge–current distribution within the past light-cone of that point. There are, of course, also advanced solutions, in which the field depends on the charge–current distributions within the *future* light-cone, but these are commonly rejected as unphysical, and textbooks often explicitly invoke considerations of causality in taking the physical solutions to be the retarded solutions.

One can compare solutions of the field equations with and without the source charges and currents. The difference between these solutions is the effect of the sources.

What has been said about classical field theories holds equally well for quantum field theories. Just as in the classical case, quantum field theories admit of formulations in which source fields or charges are treated exogenously, as classical ('c-number') terms added to the Lagrangian of the theory. One can, therefore, compare a global solution that lacks the source terms with one in which the source is zero up until some point at which it is turned on, and, just as in the classical case, take the effect of the source term on the quantum field as the difference between a global state that lacks the sources and a global state of the field with source term included. Here, again, one has a distinction analogous to retarded and advanced solutions, in the form of retarded and advanced propagators.

Thus, there is a respectable class of theories, which includes both classical and quantum field theories, that admit of formulations with exogenous source terms. Moreover, it is *essential* to certain sorts of applications that a theory admit of a formulation in which certain parameters are treated exogenously. This, as we have seen, is essential for statistical hypothesis testing, and for the employment of causal considerations.

I propose the following adequacy condition for a physical theory:

> Even if one accepts that a physical theory with aspirations to comprehensiveness ought to be capable in principle of delivering a global representation of the entire universe, leaving nothing out, any physical theory worth its salt should also be capable of *not* representing the entire universe, and should admit of a formulation with exogenous parameters.

The formulation with exogenous parameters might be limited to certain regimes of the theory's space of possibilities. It is a requirement on a comprehensive physical theory that it admit the possibility of systems that act appropriately like our experimental apparatus. It is *not*, of course, a requirement that every solution of the theory contain such things. One might, for example, have a theory of quantum gravity according to which a spacetime description is possible only if certain conditions are met, and then at a suitably coarse-grained level. The proposed condition is that the theory be capable of admitting exogenous parameters in the regimes in which we need to employ them.[3]

I am tempted to elevate this into an *a priori* condition for the possibility of physical theorising, but, in light of the chequered history of such proposals, confine myself to a more cautious formulation, echoing what Einstein said of the principle of separability: without it, 'physical thought in the sense familiar to us would not be possible'.[4] This leaves open the possibility of some hitherto-unthought-of, unfamiliar modes of physical theorising.[5]

So, for example, were someone to propose a viable superdeterministic theory that accounted for quantum phenomena via local means, *and* if an account were given of why one should take that theory seriously that was not vulnerable to SHC's charge that one is simply rejecting the usual canons of scientific inquiry because they lead to undesirable conclusions, then we would have motivation for reconsidering the proposed adequacy condition. The attitude expressed by Bell seems to be a judicious one.

> Of course it might be that these reasonable ideas about physical randomizers are just wrong—for the purpose at hand. A theory may appear in which such conspiracies inevitably occur, and these conspiracies may then seem more digestible than the non-localities of other theories. When that theory is announced I will not refuse to listen, either on methodological or other grounds. But I will not myself try to make such a theory. (Bell 1977: 83; 2004: 103)

I am sceptical that it is possible to produce a theory like that *and also* answer SHC's charge. But, like Bell, if someone were to make such a proposal, I would not refuse to consider it.

17.7 Perspective Duality and the Nature of Randomisers

Suppose we have some physical system S that is to be manipulated and experimented upon via some piece of apparatus A. On our proposal, a physical theory ought to be capable of treating the combined system $S + A$ as a physical system in its own right, subject to the dynamics of the theory, and of *also* treating just S, with the influence of the apparatus treated as exogenous. Call these two sorts of treatment the *totalising* and *limited* perspectives, respectively.

Suppose we are interested in the effect on the system of choosing between two parameter settings, a and b. Then we can treat the parameters as exogenous variables, and contrast solutions to the equations of motion for the system S with setting a, and with setting b. These will be global solutions that agree up to the time at which the parameters are set, and differ only after this.

We should be able to switch to a totalising perspective, and include the apparatus in our physical description. The two modes of description should mesh.

Exactly how this can work in the context of quantum field theory will be the subject of a sequel, but the gist of such an account will be as follows. We should be able to model the apparatus, at least schematically, within the theory. A limited perspective is obtained by tracing out the degrees of freedom of the apparatus. The presupposition motivating the microcausality axiom is that, if experimental apparatus couples to the fields under study only in a bounded spacetime region $\mathcal{O}$, its effect on the state of the fields can be modelled by a set of operators $\{\hat{M}_i\}$ belonging to $\mathfrak{A}(\mathcal{O})$. This should be a consequence of modelling the apparatus within the theory as something that couples locally to the field, and then tracing out the degrees of freedom of the apparatus.

We can compare two solutions pertaining to differing values of experimental parameter settings; the differences between the two solutions are to be counted as the effects of varying the parameters. An approach of this sort lends insight. Rafael Sorkin (1993) has shown that simply importing a measurement postulate into quantum field theory borrowed from non-relativistic quantum mechanics, to the effect that for any observable corresponding to a bounded spacetime region there is a corresponding experiment whose effect is projection onto an eigenstate of the observable, can lead to superluminal signalling. If one holds that superluminal signalling is impossible, measurements of that sort are impossible, but some principled reason is needed for *why* one should regard them as impossible. The answer is that modelling the apparatus within the theory, as coupling locally to the fields, eliminates such impossible measurements (see Papageorgiou and Fraser 2024 and references therein).

There's a potential obstacle to the meshing of the two modes of description: If the dynamics are deterministic, there are no solutions that differ only to the future of the setting event; if two solutions differ in the parameter settings, they differ at all times.

Here we can avail ourselves of another principle of physical theorising. A theory should be capable of making empirically testable predictions (of either a deterministic or probabilistic sort). As it is never possible to obtain *exact* values of all the relevant parameters, any such predictions will have to be, at least to a certain extent, robust under approximation. And, though invertible dynamics mean that there can be no pair of exact solutions that agree up to a certain time and subsequently diverge, there *can* be solutions that *approximately* agree up to a certain time and subsequently diverge to a significant extent. This means that even a slight degree of coarse-graining suffices for the possibility of two coarse-grained solutions that agree up to a certain point and thence diverge.

This is the phenomenon of sensitive dependence on initial conditions. It is a familiar fact that certain sorts of systems with deterministic dynamics are *effectively* unpredictable. That is, though an exact specification of the state at a given time uniquely determines future states, an approximation to the state at a given time may be compatible with very different states at future times. If the device responsible for the instrument settings exhibits this sort of sensitive dependence, it is possible for two solutions to agree to a high degree of approximation up to a certain point, and nevertheless contain different parameter settings, and to differ in all the consequences of these different settings.

This is, of course, true of the devices that we think of as paradigmatic randomisers. A properly flipped coin, or a die thrown, exhibits exquisite dependence on initial conditions, in such a way that information about the state of the world that is empirically accessible prior to the coin flip or dice toss is useless for prediction of the outcome.

This, then, is one way to reconcile a limited perspective, in which two solutions are compared that agree up to a point and then differ as a result of differing values of some exogenous variables, and a totalising perspective that includes the experimental apparatus in the description. If we take the totalising description to be one involving, not an exact specification of state, but some degree of coarse-graining, we can have, despite determinism at the level of fine-grained dynamics, two histories that agree up to a certain point and thence diverge.

Not every device that is used as a randomiser is of the effectively unpredictable sort. We also use pseudo-random number generators, whose output is predictable, but expected to be uncorrelated with the state of the system being experimented on. This feature is more important than unpredictability of the output of the randomiser.

Whence this lack of correlation? We might invoke a common-cause principle, to the effect that there be no correlations between the states of things that have not interacted, either directly or via some mediating third system. But this, by itself, would be empty, unless we imagine a randomiser that has somehow been kept for all eternity in isolation not only from the target system but also from anything that will subsequently be linked via a causal chain to the target system.

Isolation of this sort is an unsatisfiable, unreasonable, and unnecessary demand. Our conviction that we can produce outputs of a randomiser that are effectively uncorrelated with the state of the system being experimented upon is based, not on isolation of the randomising devices, but on the opposite, on the idea that the systems are subjected to influences that effectively efface any correlations between the device and the target system. The idea is that the dynamics of the randomising device render events in the common past of the device and the target system irrelevant to its output. They are, as J. S. Bell has put it, 'forgetting machines'.

> In such a device the complete final state perfectly determines the complete initial state—nothing is forgotten. And yet for many purposes, such a device is precisely a 'forgetting machine'. A particular output is the result of combining so many factors, of such a lengthy and complicated dynamical chain, that it is quite extraordinarily sensitive to minute variations of any one of many initial conditions. It is the familiar paradox of classical statistical mechanics that such an exquisite sensitivity to initial conditions is practically equivalent to complete forgetfulness of them. To illustrate the point, suppose that the choice between two outputs, a and a', depended on the oddness or evenness of the digit in the millionth decimal place of some input variable. Then fixing a or a' indeed fixes something about the input—i.e. whether the millionth digit is odd or even. But this peculiar piece of information is unlikely to be the vital piece for any distinctively different purpose, i.e., it is otherwise rather useless. With a physical shuffling machine, we are unable to perform the analysis to the point of saying just what peculiar feature of the input is remembered in the output. But we can quite reasonably assume that it is not relevant for other purposes. In this sense the output of such a device is indeed a sufficiently free variable for the purpose at hand. (Bell 1977: 82–3; Bell 2004: 102–3)

As Bell notes, this phenomenon is related to the process of relaxation to equilibrium in statistical mechanics. This is a process in which there is a failure of invertibility at the macroscopic scale. For many systems there is, for given values of total energy and the external parameters that define the thermodynamic state of a system, a unique equilibrium state that the system relaxes to, independent of the initial state. This is to be contrasted with invertibility at the level of microphysical dynamics, which preserves distinguishability of states. The project of explaining the tendency of systems to equilibrate can be thought of as the task of reconciling preservation of memory of the initial state at the level of microdynamics with macroscopic forgetfulness (see Myrvold 2020 for further discussion of this point).

17.8 Conclusion

The worries presented can be addressed. Moreover, it seems to me, seeing how they are to be addressed lends insight into the nature of scientific theorising. Those who have stressed that a fundamental physical theory should be capable of encompassing everything in its embrace, and should be formulable without reference to a split between system and observer, or system and apparatus, are right. But, in order for it to be applicable in anything like the way we do, in fact, apply physical theories, a theory should *also* be capable of *not* embracing everything at once in its description.

Acknowledgements

Preliminary versions of this work were presented at a symposium, *Randomness in Scientific Practice*, at the 27th Biennial Meeting of the Philosophy of Science Association, November 2021, and at the *Causalworlds* conference, at the Perimeter Institute for Theoretical Physics, September 2024. I am grateful to the discussants at both events for helpful feedback. I am particularly grateful to Marshall Abrams and Doreen Fraser for helpful discussions.

The conclusions drawn in this work, and the reasoning used to reach them, have much in common with those of Mathias Frisch (2014), who goes into these matters in much more detail than is presented here.

Endnotes

1. Recall, from §17.3, that $\mathfrak{A}(\mathcal{O})$ is the algebra of operators associated with spacetime region $\mathcal{O}$, which contains operators pertaining to observables measurable by operations in the causal closure of $\mathcal{O}$. The quasilocal algebra $\mathfrak{A}(\mathcal{M})$ is the algebra associated with the whole of Minkowski space; it's the smallest algebra that contains all the algebras associated with bounded regions of spacetime.
2. See Myrvold (2016) for an argument of this sort.
3. I am grateful to an anonymous referee for raising this point.
4. 'Ohne die Annahme einer solchen Unabhängigkeit der Existenz (des «So-Seins») der räumlich distanten Dinge voneinander, die zunächst dem Alltags-Denken entstammt, wäre physikalisches Denken in dem uns geläufigen Sinne nicht möglich' (Einstein 1948: 321). I quote from the translation of Don Howard (1985: 187).
5. Some have attributed a less cautious view to Einstein, but that's not his fault.

Bibliography

Bell, J. S. (1977), 'Free Variables and Local Causality', in *Epistemological Letters* 15: 79–84. Reprinted in *Dialectica* 39 (1985): 103–16, and in Bell (2004: 100–4).

Bell, J. S. (1990), 'Against "Measurement"', in *Physics World* 3: 33–40. Reprinted in Bell (2004: 213–31).

Bell, J. S. (2004), *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press). 2nd edition.

Einstein, A. (1948), 'Quanten-Mechanik und Wirklichkeit', in *Dialectica* 2: 320–4.

Frisch, M. (2014), *Causal Reasoning in Physics* (Cambridge University Press).

Gisin, N. (1989), 'Stochastic Quantum Dynamics and Relativity', in *Helvetica Physica* 62: 363–71.

Hitchcock, C. (2007), 'What Russell Got Right', in Price and Corry (2007: 45–65).

Howard, D. (1985), 'Einstein on Locality and Separability', in *Studies in History and Philosophy of Science* 16: 171–201.

Knox, E., and Wilson, A. (eds.) (2021), *The Routledge Companion to Philosophy of Physics* (Routledge).

Myrvold, W. C. (2016), 'Lessons of Bell's Theorem: Nonlocality, Yes; Action at a Distance, Not Necessarily', in M. Bell and S. Gao (eds.), *Quantum Nonlocality and Reality: 50 Years of Bell's Theorem* (Cambridge University Press), 238–60.

Myrvold, W. C. (2020), 'Explaining Thermodynamics: What Remains to be Done?', in V. Allori (ed.), *Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature* (World Scientific), 113–43.

Myrvold, W. C. (2021), 'Relativistic Constraints on Interpretations of Quantum Mechanics', in Knox and Wilson (2021: 99–121).

Myrvold, W., Genovese, M., and Shimony, A. (2024), 'Bell's Theorem', in E. N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University). Spring 2024 edition. https://plato.stanford.edu/archives/spr2024/entries/bell-theorem/

Norton, J. D. (2007), 'Causation as Folk Science', in Price and Corry (2007: 11–44).

Papageorgiou, M., and Fraser, D. (2024), 'Eliminating the "Impossible": Recent Progress on Local Measurement Theory for Quantum Field Theory', in *Foundations of Physics* 54: 26.

Pearl, J. (2000), *Causality: Models, Reasoning, and Inference* (Cambridge University Press).

Peirce, C. S. (1931), *The Collected Papers of Charles Sanders Peirce, Volume 1: Principles of Philosophy* (Harvard University Press).

Price, H., and Corry, R. (eds.) (2007), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited* (Clarendon Press).

Ruetsche, L. (2011), *Interpreting Quantum Theories* (Oxford University Press).

Ruetsche, L. (2021), 'Locality in (Axiomatic) Quantum Field Theory: A Minority Report', in Knox and Wilson (2021: 311–22).

Russell, B. (1913), 'On the Notion of Cause', in *Proceedings of the Aristotelian Society, New Series* 13: 1–26.

Shalm, L. K., Meyer-Scott, E., Christensen, B. G., Bierhorst, P., Wayne, M. A., Stevens, M. J., Gerrits, T., Glancy, S., Hamel, D. R., et al. (2015), 'Strong Loophole-Free Test of Local Realism', in *Physical Review Letters* 115: 250402.

Shimony, A. (1993), *Search for a Naturalistic Worldview, Volume II: Natural Science and Metaphysics* (Cambridge University Press).

Shimony, A., Horne, M. A., and Clauser, J. F. (1976), 'Comment on "The Theory of Local Beables"', in *Epistemological Letters* 13: 1–8. Reprinted in *Dialectica* 39 (1985): 97–102, and in Shimony (1993: 163–7).

Simon, C., Bužek, V., and Gisin, N. (2001), 'No-Signalling Condition and Quantum Dynamics', in *Physical Review Letters* 87: 170405.

Sorkin, R. D. (1993), 'Impossible Measurements on Quantum Fields', in B. L. Hu and T. A. Jacobson (eds.), *Directions in General Relativity: Proceedings of the 1993 International Symposium, Maryland, Volume 2: Papers in Honor of Dieter Brill* (Cambridge University Press), 293–305.

Woodward, J. (2007), 'Causation with a Human Face', in Price and Corry (2007: 66–105).

Index

abstractness 7, 26, 32n22, 94–107, 151, 274. *See also* generality
action 75, 160–62, 250–51, 254–6. *See also* Yang–Mills theory
  at a distance 3, 113, 115–16, 119, 121, 126n14, 126n18, 195, 283, 322
adaptation 223. *See also* response, adaptive; selection, adaptive
agency 216, 219–22, 226. *See also* response, adaptive, agency-enabling
algebraic quantum field theory (AQFT) 30–31n4
Araki–Yanase theorem 288
area law/rule 115
asymmetry, temporal/time 251, 320. *See also* symmetry
autonomy 9, 39, 46, 52, 73–6, 84, 194, 202–5, 247–8, 252–3, 257, 258n7. *See also* dynamics, autonomous; model, autonomous; non-autonomy; subsystem, autonomous/independent; system, autonomous
  absolute (A-autonomy) 203–5
  parameterised (P-autonomy) 204–5, 208–9

BBGKY hierarchy. *See* Bogoliubov–Born–Green–Kirkwood–Yvon hierarchy
Belinski–Khalatnikov–Lifschitz conjecture 253
Bell inequalities/Bell's inequality 58, 324
biology 1, 217–18, 221, 225. *See also* hierarchy, biological; organism; system, biological
black hole 11, 19, 65, 87–8, 169n3, 290
  information loss 88, 169n3, 265, 290
Bogoliubov–Born–Green–Kirkwood–Yvon (BBGKY) hierarchy 91n2
Bohmian mechanics. *See* mechanics, Bohmian
Born rule 21, 84, 265
boundary 8, 59, 147–66, 195–6, 203, 216, 219, 222, 224, 293, 306–7
  ~ reference frame 166
  spacetime 148–9, 159
Brownian motion 72, 193–4, 198

Caldeira–Leggett equation 208, 210
Cantor's theorem 308
causality 309, 315n21, 318, 323, 326. *See also* microcausality; quantum causality
  relativistic 46, 320–21, 323–4
Choi–Effros theorem 299
Choi–Jamiołkowski isomorphism 297
clock 148, 151, 153–4, 157, 163, 165–6, 169n2, 187–8, 314n11
  quantum 163
closed system 3, 8–11, 18–19, 26, 29, 40, 47, 49, 57, 59–63, 65–6, 74, 83, 86–7, 89, 105–7, 113, 118, 124, 132, 149, 176, 193–201, 205, 209, 211, 219, 222, 234, 237–8, 243, 247–8, 252, 258, 283–4, 286, 290, 295, 297, 311. *See also* open system
closed systems view 5–8, 19–20, 26, 29–30, 32n25, 38, 44, 49, 51–3, 57, 61–5, 87, 101–2, 113, 130, 132–3, 135–7, 144, 239, 241–2, 283–4, 290. *See also* open systems view
  temporally 44
closure 10–11, 38, 40, 43–4, 47, 50, 53–4, 223, 226, 247–58, 293, 305–11, 320, 331n1. *See also* closed system
collapse 23, 31n10, 81, 84, 86, 212n8, 225, 261–2, 264, 268–72, 278, 283, 290, 317
Collins–Popescu conjecture 286–90
commutativity 30n4. *See also* noncommutativity
completeness 6, 17, 20, 22–4, 26, 29, 31n10, 47, 50, 71, 233, 238, 252–3, 256–7, 307, 319
complexity 9, 66, 216–17, 244, 302–3
composition 33n28, 191n2, 296. *See also* decomposition
conservation (of some quantity) 9, 11, 194, 196–7, 200–1, 205–7, 209, 212n8, 240–41, 247–8, 253–4, 283–90. *See also* non-conservation
constructivism 312
context 9–10, 21–2, 24–6, 28, 32n15, 174–8, 186–90, 190n1, 216–19, 222–3, 225, 314n20
contextuality 157, 177, 189
contraction 85–6, 296–8. *See also* dilation
Corollary VI (of Newton's *Philosophia Naturalis Principia Mathematica*) 3, 199

cosmology 10, 19, 71, 195, 205, 210, 212, 245, 247, 253–8. *See also* quantum cosmology
coupling (of/between systems) 5, 25, 76, 79, 81, 85, 113, 132, 149, 161–3, 185, 203, 288, 328. *See also* decoupling
covariance 149–50, 152, 161–2, 164–8, 169n4, 320, 323–4
CP. *See* map (dynamical), completely positive
CP (charge conjugation parity) violation 80

de Broglie–Bohm pilot-wave theory 317
decoherence, quantum 4–5, 7, 27, 47, 59–61, 86, 179, 190n1, 197, 201, 206–7, 265
decomposition 29, 38, 45–6, 49, 51, 58, 61, 74, 161, 297. *See also* composition
decoupling 161. *See also* coupling
density matrix 4, 106, 206, 268, 272–3, 275
  realism (DMR) 10–11, 260–79, 280n1 (*see also* wave function realism)
  universal/of the universe 263–6, 269, 271, 275–6, 280n1
density operator 17–18, 21, 26–9, 30n2, 33n28, 45, 48–9, 63–4, 75, 87–8, 103, 206–8, 243, 245, 295, 297, 299
  mass-~ 264
dependence 38, 41–2, 76, 120, 125, 147, 154, 156, 157, 166, 194, 198, 204–6, 220, 235, 247, 252–3, 258, 283, 299, 326, 329
determinism 54n1, 219, 225, 256, 279, 319, 322, 329. *See also* indeterminism; superdeterminism
development 216, 219, 223
dilation 7, 78–9, 83, 85–90, 298. *See also* contraction
  theorem (*see* Stinespring dilation theorem)
dissipation 4–5, 206, 213n11
DMR. *See* density matrix realism
DNA (deoxyribonucleic acid) 223, 310
dynamics 7, 18–21, 29–30, 31n9, 32n25, 40, 42, 46, 52, 72–4, 77–81, 83–90, 97, 105, 147–50, 155–6, 158, 162–3, 165, 167, 187–8, 193, 197–212, 213n10–11, 226n2, 241, 254–5, 261, 266–8, 270–71, 274, 276, 295–6, 317, 321, 328–30. *See also* electrodynamics, classical
  autonomous 52, 75–6, 84–5, 88–90, 91n2, 202, 208
  coarse-grained 202, 258n3
  non-unitary 7, 31n7, 72, 73, 79, 84, 87–8, 90, 206–8, 212
  unitary 7, 31n10, 61, 72–4, 77–9, 83–4, 86–8, 90, 91n4, 206–8

earth 113–18, 122–3, 125n2, 126n8, 126n18, 127n21, 200, 203, 306
Ehrenfest relation 208, 211
Einstein–Hilbert formulation of gravity 160
Einstein's equations 156, 256
electrodynamics, classical 138, 142
electromagnetic theory 134–5, 141, 169, 319, 326
embeddability 194, 198, 210–11
embodiment (of an observer/agent) 147, 149, 167
emergence/emergent structure 9, 47, 62, 101, 183–5, 194, 202, 216–26, 261, 265, 304, 310, 314n20
emission 131, 134, 136, 139–40, 144
energy 206, 216, 220, 247–9, 251–4, 256–7, 258n6, 286–9, 301, 330
  conservation (*see* conservation)
entanglement, quantum 4, 7, 31n4, 43, 51, 58–9, 62, 81, 157, 165, 233, 239–41, 262, 266, 273, 277, 286–90, 296, 302. *See also* Collins–Popescu conjecture; separability
entropy 131–6, 206, 220
  gravitational 157
  informational 206
environment 1–5, 7, 9–11, 17–19, 27–8, 31n7, 40, 59, 61, 63, 66, 72–3, 76–7, 80–81, 83–90, 105–6, 113, 132, 135–6, 150, 190n1, 193–200, 203, 205–7, 209–11, 216, 220–23, 234, 238, 240, 249–50, 269–75, 284, 299, 304, 317–18. *See also* interaction, system–environment
epistemology 90, 168, 312, 314n13. *See also* onto-epistemology
EPR (Einstein–Podolsky–Rosen) 27, 61, 183
equilibration 76, 88–90, 330
equivalence, empirical 266–8, 272–4, 278
equivariance 263, 268, 270
Euler–Lagrange equations 153, 250
evolution (of a system) 4, 18–19, 21, 29–30, 31n10, 32n25, 38, 40–42, 45, 47–8, 50–53, 64, 72–3, 76–7, 85, 87, 106, 132, 135, 152–3, 178, 187, 194–6, 200–201, 203, 241, 251–2, 256–8, 264, 266, 270, 284, 296, 328–9. *See also* dynamics
  Darwinian 10, 221, 223
  non-unitary 4, 18, 61, 87, 198, 243, 271, 290
  unitary 4–5, 9, 47, 50–51, 63, 90, 178, 186–8, 194, 197–8, 206–7, 209–10, 213n10, 295–7
explanation 78, 83, 89, 99, 234–5, 237, 239, 258, 272, 278–9
explosion, principle of 306
extensibility, indefinite 10, 233–4, 243–5, 300

falsifiability 312
feedback circuit/loop 221–3
Fock space 108n3
frame 31n15, 147–68
  idealised 147, 158, 165, 167
  physical 147, 149, 151, 153–68
  reference ~ (*see* reference frame)
framework 1–2, 4–7, 9, 17–30, 31n5, 42–5, 57, 64, 66, 81, 87, 94–107, 107n2, 132, 137, 177, 183, 186, 189, 234–7, 239, 241–2, 283–4. *See also* theory, Type A
  fundamental (*see* fundamentality)
  –theory–model distinction 94–5, 98
Frauchiger–Renner thought experiment 31n12, 167
free will assumption 324
Friedmann equations 255
fundamentality 45, 64, 66, 94, 100–104, 106–7, 132, 137, 234–7, 241

gas 194, 201
  laws 194
  ~es, kinetic theory of 133
gauge 147, 149, 152–6, 159–66, 169n5. *See also* symmetry, gauge; theory, gauge
gene 220, 223
generality 87, 94–107. *See also* abstractness
general quantum theory of open systems (GT) 18–19, 29–30, 44–6, 51, 62–4, 66, 91n3, 95–6, 101–4, 106–7, 238, 241. *See also* standard quantum theory
geodesic 154, 256–7
geometry 71, 148, 151, 156–7, 161, 185, 212, 247, 253, 300
Ghirardi–Rimini–Weber (GRW) theory 23, 43, 50, 81, 263–4, 268
Gleason's theorem 17, 30n2–3
GR. *See* relativity, general
gravitation 3, 8, 65, 73, 79–80, 118, 122–5, 126n8, 126n18–19, 147–50, 152, 154–9, 162–8, 196, 198–200, 203–4, 318
  quantum (*see* quantum gravity)
  universal 114–15, 118–19
Gribov ambiguity/obstruction 155–6
GRW theory. *See* Ghirardi–Rimini–Weber theory
GT. *See* general quantum theory of open systems

Halley's comet 114, 119
halting problem 302, 308
Hamiltonian 5, 18, 32n25, 62, 65–6, 72, 74–5, 77, 97, 159, 169n4, 197, 201, 206–7, 247–57, 258n4, 258n7, 266, 279, 284, 286, 288, 296
  contact ~ 201, 250–53, 255–8
  symplectic 159, 169n4, 208, 247–8, 250–58
Hamilton's equations 208, 252, 254
Hamilton's principle 248
harmonic oscillator. *See* oscillator
Hartle–Hawking state 265
Hawking–Penrose theorems 256
Herglotz–Lagrange equations 250, 251, 255
Hertzian
  oscillator (*see* oscillator)
  resonator 133, 140
hierarchy 9, 77, 91n2, 216–18, 309–10
  biological 217–18
  tangled 309–10
Hilbert space 18, 25–6, 30n2, 60, 64, 72, 74, 157, 159, 169n3, 174, 181, 186–7, 206, 266, 272, 275, 277–8, 299, 314n16
Hill–Brown lunar theory 113, 117
holism 40, 60, 169n3, 233, 238–9
homeostasis 216, 220–21
Hubble
  friction 255
  parameter/expansion 254–6, 258, 314n12

idealisation 5, 8, 11, 19, 65, 73, 77, 79, 84, 95, 98, 101–2, 114–15, 121, 147–9, 152, 158–9, 162, 165–7, 175, 195, 200, 256, 298
inclosure schema (Schema T) 306
incompleteness theorem (first) 307
independence 6, 22, 24, 26–7, 38–41, 48, 50, 57, 66, 94, 132–3, 140, 153, 156, 164, 196, 202–3, 205, 247, 250–51, 253, 258, 270, 304, 320, 324, 330
  measurement ~ 174–6, 183, 324–5
  statistical 324
indeterminism 20, 41–2, 153–4, 301. *See also* determinism
indistinguishability 5, 21, 61–2, 272–3
infinity 9, 11, 107–8n2, 144n4, 149, 152, 158–62, 165, 167, 175, 177, 186, 188–90, 244–5, 254, 256–7, 293–313, 314n11–12, 314n16–17
information 2–4, 23, 33n28, 38, 41, 45–6, 66, 85, 88, 132, 148, 178–82, 187–8, 196, 216–17, 220, 222–4, 239, 249–50, 274, 277, 280n3, 283, 290, 300, 302, 314n20, 329–30
  loss 88, 290 (*see also* black hole information loss)
initial projection hypothesis 262, 265
integrability, breakdown of 203
integration, interlevel 222–3
interaction 2, 4–5, 7, 18, 21–2, 24–6, 32n15, 32n22, 32n25, 39–40, 50–51, 58, 66, 74–6, 78–81, 84–6, 90, 95, 98, 130, 134,

136, 139–41, 143, 147, 151, 157, 166, 178–81, 199–201, 206, 210, 224, 240–41, 244, 284, 286–8, 294–5, 303, 319, 325
gravitational 8, 113, 122–5, 126n18, 127n21, 149, 163, 198–9
measurement ~ 178–9, 284, 286, 288, 294, 317
system–environment 1–2, 5, 66, 81, 85–6, 88, 90, 105, 136, 193, 198, 200, 211, 216, 221–3, 240, 250, 294–5, 299
interpretation(s) 71, 96, 100, 102, 104, 106–7, 131, 162–3, 184, 199, 208, 240–41, 295, 297, 324
of quantum mechanics 2, 9–11, 20, 22–9, 31n10, 31n13, 32n15, 32n18–19, 32n24, 39, 43, 47–51, 62, 64–5, 71, 102–4, 167, 176–9, 182–3, 261–3, 265–7, 277, 279, 280n3, 283–90
pristine 91n1
intersubjectivity 167–8, 179
invariance 62, 75–6, 91n4, 149–50, 152–6, 158–66, 168, 169n3, 178, 183, 189, 206, 247–8, 254, 320
invertibility 319, 322, 329–30
failure of 330
irreversibility 1, 61, 74, 83, 88–9, 91, 187, 296. *See also* reversible; time, arrow of
isolation 1, 4–5, 7–9, 11, 17–19, 31n6, 39–40, 43, 63, 66, 71–4, 77–8, 83–4, 89–90, 113–16, 121–4, 134, 177, 189, 193–4, 196–201, 203, 205–6, 209–10, 212n4, 224, 245, 283, 294–5, 299, 329–30. *See also* non-isolation; system, isolated

Jupiter 3, 114–15, 119, 124, 125n4, 126n16–18
moons of 3, 8, 113–16, 125–6n3–4, 126n14
~/Saturn great inequality 114, 124

Kirchhoff–Clausius law 135
Kirchhoff's law 133, 140
Klein–Gordon equation(s) 155, 255
KMS condition/state. *See* Kubo–Martin–Schwinger condition/state
knowledge 4, 25, 66, 174, 178–81, 221, 242, 250, 285, 294–5, 303–4, 313, 313n3
accessibility of 181
acquisition 294, 313n3
lack of 262
Kochen–Specker no-go theorem 9, 176, 178
Kraus representation theorem 31n4
Kubo–Martin–Schwinger (KMS) condition/state 105, 108n6

Lagrangian 96–7, 103, 159, 247–8, 250–51, 326
Langevin equation 83
Laplace's Demon 219–20
last scattering surface (LSS) 225–6
law 11, 21, 25, 41–2, 45, 61, 65, 78, 86, 113, 131, 133–6, 138–40, 143, 193–4, 202, 216, 219–20, 224, 234, 240–41, 247–8, 253, 262, 264–7, 274, 278–9, 283–90, 317–20, 323–4
fundamental 83, 239–41, 276, 279, 280n6
of motion (*see* Newton's laws of motion)
Lawvere's fixed-point theorem 308–9, 311
Liar (paradox/sentence) 307–8, 315n22
Lindblad equation 4–5, 40, 42, 44, 65, 77, 87, 106, 207–8, 210, 275–6, 280n6, 296
Liouville's theorem 248–9
Liouville–von Neumann equation 28, 65. *See also* von Neumann equation
locality 61, 169n3, 240, 266, 274. *See also* nonlocality
logic 32n26, 33n29, 137, 141, 209, 221, 226, 234, 244, 305–6, 308–10
Lorentz covariance 320
LSS. *See* last scattering surface
lunar precession (problem). *See* precession, lunar

machine 307, 330. *See also* Turing machine
forgetting ~ 330
nomological 212n5, 225
map (dynamical) 18, 31n7, 155, 160, 164, 205–8, 210, 297, 302
completely positive 208, 210, 295–7, 299
not completely positive 241
Markov process 207
master equation 82, 207–11, 213n11, 280n6. *See also* Caldeira–Leggett equation; Lindblad equation
matter 1, 23, 39, 43, 71, 80, 82, 108n4, 115, 118, 122, 132, 139, 143, 151, 156, 162, 193, 196, 200, 216, 242, 253–4, 256, 258n6
condensed 82, 86, 90, 98, 107n1, 108n4, 304 (*see also* physics, condensed-matter)
prime 305
Maxwell's equations 142–3, 219
measure 17, 194, 197, 201, 212, 248, 251, 253, 255–8, 258n2, 258n7, 267–8, 273–4
compression 201, 210
of existence 285
expansion 201
measurement 5, 11, 18, 21, 25–8, 30n2, 31n9, 43, 47, 50, 52–3, 64, 117, 119, 161, 174–80, 182–9, 191n3, 198, 212n3, 268–71, 283–90, 294, 297, 299, 321–2, 325, 328

measurement (*Continued*)
independence (*see* independence, measurement)
~ interaction (*see* interaction, measurement)
positive operator-valued (POVM) 297
problem 5, 20–22, 26, 31n9, 31n12–13, 50, 64, 83–4, 178–9, 186, 248, 260–61, 321
of the universe 26, 32n24
weak 287–8
mechanics 1, 65, 117, 154, 199, 202, 204–5, 250
Bohmian 22–3, 31n10, 64, 176, 261–3, 267–8, 275, 280n3
classical 139–41, 175–6, 196–7, 204–5, 236, 266, 272, 301, 330
contact 250
particle 74, 90, 154 (*see also* physics, particle)
quantum (QM) 1–2, 4, 9–10, 17, 22–3, 32n17–19, 38–9, 42–51, 57, 60–67, 72, 78–91, 91n4, 96, 108n2, 129, 166, 174–90, 191n2–3, 206, 210, 212n8, 260–80, 283, 301, 311, 317, 320–22, 325, 328
statistical (*see* statistical mechanics/physics)
Mercury perihelion (problem) 114, 125, 126n18
metabolism 216, 220–21, 223
metaphysics 6–7, 10–11, 18–19, 22, 29, 31n15, 38, 40, 44–5, 53, 66, 78–80, 82, 84, 86, 94, 96, 100–104, 106–7, 132, 199, 216, 233–9, 242–5
of physics 233, 242, 245
methodology, scientific 1, 7–8, 19, 29–30, 32n25, 39–42, 47, 78–83, 86–7, 90, 98, 100–101, 104–7, 113–14, 125n1, 129–30, 132, 135, 147, 211, 235, 312, 324–5, 327
microcausality 320–24, 328
mixed state. *See* state, mixed; mixture
mixture (of pure states) 28, 270–71, 277
improper 4, 17, 63–4
proper 4, 18, 21, 27, 63–4
mode (for application of a concept/term)
formal 195, 197–202, 205, 209, 212n1–2
material 195–200, 202, 205–6, 209–10, 212n1–2, 212n7
model 5, 7–8, 10, 29, 42–4, 59, 65, 71, 80, 81, 94–107, 108n4, 113–17, 119, 121–5, 126n18–19, 127n21, 133, 140–41, 147–54, 156–9, 161–7, 169n5, 193–5, 197–212, 212n7, 222, 235–7, 244, 249–50, 253–4, 258, 258n3, 276, 322. *See also* framework–theory–model distinction
autonomous 194, 202–5, 212n7, 249
cosmological 71, 194, 210–12, 256, 258
standard ~ (of particle physics) 80–81, 87, 90, 95, 97, 108n4
momentum 58, 156–7, 183–5, 187, 189, 219, 240, 249, 270–71, 284–90
monad, Leibnizian 315n24
monism 233, 237–9, 242, 313n1, 313n8. *See also* pluralism
moon 8, 113–14, 115–20, 126n7–8, 127n21, 200, 203, 300. *See also* Jupiter, moons of; precession, lunar

Nakajima–Zwanzig equation 75, 88
narratability 38, 45–7
naturalism 74, 90, 181
Navier–Stokes equation 83
Neptune 114, 124
Newton's cradle 199
Newton's laws of motion 3, 114–16, 120–22, 199–200, 219, 288, 318
Noether's theorems 153, 240–41, 248, 284
nomic integrity, argument from 239–41
noncommutativity 30n4, 157, 320. *See also* commutativity
non-conservation 196, 200–201, 203–4, 253, 287, 289. *See also* conservation
of conserved quantities 285, 289
of energy 247
non-equilibrium thermodynamics. *See* thermodynamic(s), non-equilibrium
non-isolation 72, 199–200, 210. *See also* isolation
nonlocality 2, 150, 153–5, 159, 169n3, 177, 324. *See also* locality
nonnegativity 295–6
non-unitarity 4–5, 7, 11, 18, 29, 31n7, 47, 51, 60–61, 63–4, 72–3, 77–81, 84–90, 198, 206–8, 210, 212, 212n2, 238, 243, 271, 275, 290. *See also* unitarity
fundamental 47, 80
no-signalling condition 321
number 151, 206, 295, 298–302, 307–8, 310–11, 313n8, 314n12, 314n15–17, 326
nutation (of the earth) 116–17, 126n8

objectivity 8, 24, 28–9, 32n19, 32n21, 168, 174–8, 183, 185, 189, 233, 242, 260–62, 274–5, 311. *See also* intersubjectivity
observable (value/quantity) 17, 20–22, 25–6, 30n4, 31n11, 58–9, 61–2, 64–5, 143, 149–50, 152–8, 162–3, 165, 167–8, 169n5, 183, 248, 252, 255–6, 268, 285, 314n12, 320–21, 325, 328, 331n1

observer/agent 8, 22–4, 26, 31–2n15, 32n20, 42, 47, 50–51, 147–53, 158, 163, 165, 167–8, 169n1, 178–80, 317, 331. *See also* system–observer split
Occam's razor 11, 278, 302. *See also* parsimony, modal
Olbers' paradox 226n1
onto-epistemology 313n3, 314n13
ontological models framework 23–6
ontology 6, 20, 26, 29–30, 31n13, 38–9, 41, 45, 47–51, 60–62, 66, 99, 101–4, 107, 164, 195, 199, 201, 234–6, 238–9, 243, 263–6, 268, 278, 294, 297, 304, 313n2, 314n13
open–closed distinction 9, 42, 149, 193–6, 198, 205, 212n2, 238. *See also* unitary–nonunitary distinction
opening-up 8, 113–17, 119, 123
open question argument 237, 241–2
open system 1–2, 4–11, 17–20, 24–6, 28–30, 32n25, 40, 42, 44, 46–7, 49, 51, 57, 59–60, 63–4, 66, 83–7, 91n3, 94–5, 102, 104–7, 113–14, 117, 119–20, 122, 124–5, 132, 136, 193–4, 196, 198, 200–201, 203, 205, 208–10, 213n10, 216–21, 223–6, 233, 238, 240–41, 243, 247, 249, 253, 255, 274–6, 280n6, 284–5, 289–90, 295, 297, 311, 313. *See also* closed system; quantum system, open
  fundamentality of 94, 102, 104, 233, 237, 242
  temporally 41, 43–4, 49 (*see also* open systems view, temporally)
open systems view 5–8, 10–11, 19, 38, 40–41, 43–4, 53, 62–3, 65, 74, 78, 83–90, 91n3, 101–2, 104, 106–7, 113–14, 117, 124, 130, 132–4, 136–7, 144, 210, 233, 238–45, 275–6, 311. *See also* closed systems view
  temporally (TOSV) 6, 38–54
organism 1, 10, 193, 216–23
oscillator 62, 65, 133–7, 139, 193, 199, 204–5, 209, 249–53, 256, 326

Page–Wootters formalism/mechanism 165, 187, 289
parameter, exogenous 194, 198, 203–5, 317, 326–30. *See also* variable, external/exogenous
parsimony, modal 278. *See also* Occam's razor
particle mechanics. *See* mechanics, particle
past hypothesis 258, 265, 279, 280n3
perspectivalism 8, 32n21, 147, 149, 163–8
perspective 2, 11, 61, 105, 148, 163–8, 175–7, 183, 189, 211, 260, 318, 328–9
  limited 318, 328–9
  observer/agent-based 148, 163, 165, 167–8 (*see also* observer/agent)
  totalising 318, 328–9
perturbation 8, 82, 117, 119, 124–5, 126n17–18, 198, 221, 293, 298
phase transition 101, 298
phenomenology 74, 80–82, 86, 95, 98
Philosophiae Naturalis Principia Mathematica (Isaac Newton) 114, 119–23, 126n9
physics 1–2, 5–8, 19–20, 23, 27–9, 38, 41–2, 49, 57, 63, 65–6, 71–4, 77–88, 90, 94–9, 102, 105–7, 129, 131, 136, 145n7, 147, 149–50, 156–8, 162–8, 169n1, 174–6, 178, 190, 193–4, 198, 200, 205, 221, 225–6, 226n2, 233–5, 238–43, 253, 262, 283, 293–4, 298–9, 301–3, 312, 313n7, 314n16–17, 318–21, 323
  condensed-matter 82, 86, 90, 98, 107n1, 108n4, 304
  gravitational 8, 147–50, 152, 154, 156–9, 162–8
  non-relativistic 18, 48–9, 74, 79–81, 90, 94–9, 103, 105, 107n1, 108n2, 108n6, 154, 163, 244, 320, 322, 328
  particle 80, 95, 98, 107n1, 108n4, 219, 237 (*see also* mechanics, particle)
  relativistic 18, 21, 30n4, 43, 45–6, 48–50, 53, 62, 80, 94–7, 103, 105, 107n1, 108n2, 108n6, 125, 127n21, 319–21, 323–4
  statistical (*see* statistical mechanics/physics)
Planck scale 81
pluralism 66, 78, 90, 233, 237–9. *See also* monism
plurality 11, 83, 293, 305–6, 311–12
pointillisme 39, 43–4. *See also* reductionism
possibility 305, 327. *See also* world, possible
  epistemic 233, 242
  mathematical 242
  objective 233, 242, 278
potential 151, 160, 162, 197, 204–6, 254, 318
  symplectic 169n4, 251–2
  ~ well 205
POVM. *See* measurement, positive operator-valued
precession
  of equinoxes 113–14, 117, 119
  lunar 116, 118
    problem 113, 124
  Mercury perihelion 125
prediction 5, 17–18, 41, 120, 130, 141, 163, 175, 179, 222, 254, 260, 268, 270, 276, 278, 283, 329. *See also* processing, predictive

Principia. *See* *Philosophiae Naturalis Principia Mathematica*
probability 4, 17–18, 20–21, 24–8, 31n13, 32n25, 32n27, 50–51, 84, 133, 139, 174–7, 180, 182, 184–8, 190n1, 196–7, 206–7, 225, 264–5, 267–8, 270, 272, 275–6, 279, 285, 296, 314n16, 320–21
  conservation 87, 206–8, 248
  illusion of 285
  interpretation of 32n19, 32n21, 248
process, irreversible 1, 88, 187, 296. *See also* irreversible
processing, predictive 216, 219–20, 222
property
  absolute 2, 58, 98, 193–4, 196, 199–200, 203–5, 209, 212, 288
  scale-relative 193–4, 198–201 (*see also* scale-relativity/dependence)
protectorate
  classical 217
  quantum 217
provability 307–8
pure state. *See* purification; state, pure
purification (of a mixed state) 28, 33n28, 277, 293, 296–7, 304
Pusey–Barrett–Rudolph theorem 23

QBism 23–4, 31n15, 32n19, 32n21, 50, 177. *See also* quantum Bayesianism
QCD. *See* quantum chromodynamics
QED. *See* quantum electrodynamics
QFT. *See* quantum field theory
quantisation 80, 130, 137–44, 208–9, 211–12
quantum
  Bayesianism 177 (*see also* QBism)
  causality 2
  channel 295–6
  chromodynamics (QCD) 82, 95, 97–9
  clock (*see* clock, quantum)
  coherence 4, 197
  computation 85
  cosmology 153–4, 162–3, 209–11, 213n11, 260, 262, 265
  decoherence (*see* decoherence, quantum)
  dynamics 73, 78, 82, 90, 207, 209, 211, 226n2, 321
  electrodynamics (QED) 80–81, 90, 95, 97–9
  entanglement (*see* entanglement, quantum)
  equilibrium distribution 263, 268
  field 31n4, 326
  field theory (QFT) 7, 30n4, 48–9, 62, 74, 77, 88, 94, 198, 320, 322, 324–6, 328
  game 298
  gate, unitary 187
  gravity 8, 54n3, 80–81, 147, 149, 151, 153–4, 156, 158, 162–3, 165, 167–8, 260, 276, 327 (*see also* gravitation)
  holism 238
  information theory 201, 302
  interference 82
  magic square 298
  measurement (*see* measurement)
  mechanics (*see* mechanics, quantum)
  Mentaculus 278 (*see also* Wentaculus)
  nonlocality (*see* nonlocality)
  optics 80–81, 286
  pluralism 78
  protectorate (*see* protectorate)
  state 4–5, 10–11, 20–22, 24, 47, 49–50, 58, 60, 72, 75, 90, 167, 174, 177–82, 186, 190, 206, 260–62, 265–7, 269–75, 277–9, 283–7, 290, 297, 321–2
  statistical mechanics (*see* statistical mechanics/physics, quantum)
  system (*see* system, quantum(-mechanical))
  uncertainty (*see* uncertainty)
qubit 58, 179–80, 186, 191n2, 313n5

radiation 8, 72, 81, 129–44, 218, 285, 290, 314n12
  theory 30, 129–44
randomness 264, 283, 285, 318, 321, 324–5, 327–30
Rayleigh–Jeans law 131
realism 10, 22, 24, 32n21, 49, 66, 145n6, 164, 233, 236–7, 242, 260–62, 276, 279
  density matrix (*see* density matrix realism)
  spacetime state 48–50, 54n3
  wave function (*see* wave function realism)
reality 4, 6, 10, 20, 24, 26, 29, 38, 40–41, 46, 52, 65, 151, 183, 236, 243–5, 261–2, 293–4, 313n1, 313n8, 314n21, 317, 319
  as a plurality 311–12
  as a unity 311
  as a whole 71, 233, 243, 318
reasoning 116, 118–19, 177, 186, 194, 211, 312
  causal 89, 318
  strategies, complementary 8, 129–44
reduction 1, 202, 234, 237
reductionism 7, 38–41, 44, 46–8, 51, 73–4, 78, 80, 82–4, 86, 88–91, 297, 303–4, 313n8, 314n19. *See also* pointillisme
reference frame 8, 19, 43, 48–9, 53, 87, 147–52, 154–68, 175–6, 189–90, 288
  ideal(ised) 147, 165
  physical 8, 147, 149–51, 155–68
  quantum 8, 19, 87, 147, 157–8, 162, 164–5, 167, 190

relativity 7, 50, 57–60, 63, 122, 240, 305
general (GR) 71, 125, 127n21, 152, 154–5, 159, 164, 169n3, 198, 226, 236, 247, 253–4, 256, 258n6
special 218, 320, 323
reservoir 208–11
resources 20, 26, 29, 88, 114, 175, 179–81, 186, 189, 191n3, 237
finite 9, 175, 177, 186–7, 189–90
infinite 9, 177, 186, 188–90
practically 175, 189
response 82, 182, 219, 221–3
adaptive (predictive) 216, 219
agency-enabling 219
Rule 1 (of Newton's *Principia*) 118, 123, 126n9
Rule 2 (of Newton's *Principia*) 118, 123, 126n9
Rule 3 (of Newton's *Principia*) 114, 118–19, 126n10
Rule 4 (of Newton's *Principia*) 114, 119–20, 126n12
Russell's paradox 308

scale (factor) 3, 19, 26, 71–2, 74–5, 81, 88, 90, 98, 100, 108n4, 193, 199–203, 210–11, 217, 221–2, 226, 245, 254–5, 257, 311, 330. *See also* timescale
numerical 201
Planck (*see* Planck scale)
-relativity/dependence 71, 193–4, 198–201, 203
separation of ~s 75, 90, 91n2
scaling 254–5, 303. *See also* symmetry, scaling
Schema T. *See* inclosure schema
Schrödinger equation 4, 21, 44, 65–6, 72, 88, 207, 263, 267
Schrödinger's cat 5
selection 21, 220, 222
adaptive 223 (*see also* adaptation; response, adaptive (predictive))
natural 223
self 309–10
-organisation 193
-reference 307–9
-transparency 309
separability 5, 41, 43–4, 46, 176, 265–6, 288, 327. *See also* entanglement, quantum
signalling, superluminal 320, 328
similarity vector/dynamical similarity 252
simultaneity 98, 154
singularity 203–4, 256–8
slice, spatial/spacetime 48–9, 253–4, 256, 319. *See also* time slice
society 10, 218, 222, 224–5
solar system 8, 113–14, 118, 120–5, 126n14, 126n18, 199, 201, 218, 273
space 39, 48–9, 52, 57, 98–100, 102, 122, 143, 151–4, 165, 169n1, 174–5, 179–80, 183–5, 188–90, 203, 244, 263–5, 276, 284, 300–301, 305, 314n12, 322, 326
emergent (notion of) 184–5, 261
infinite 300
spacetime 6, 43, 47–53, 57, 97–9, 122, 147–67, 169n1, 188–9, 244–5, 247, 256–7, 264–6, 319–20, 322–3, 325–8, 331n1
state realism 48–50, 54n3, 261
spectral theorem 296
spin 9, 27, 58–60, 62, 98, 174–90, 191n3, 275, 287–8
network 9, 175, 177, 183–90, 191n3
evolving 187
spinthariscope 285
ST. *See* standard quantum theory
standard model of particle physics. *See* model, standard (of particle physics)
standard quantum theory/mechanics (ST) 4, 6–7, 9, 18–21, 23, 26, 28–30, 31n9–10, 32n24–5, 44–6, 51, 57, 60–62, 64, 66, 95–6, 101–4, 107, 108n2, 183, 241, 325. *See also* general quantum theory of open systems
state 4–5, 17, 20–22, 24–9, 30n2, 30n4, 31n10, 31n14, 32n25, 41–2, 45–51, 58, 60–61, 63–4, 66, 72–9, 82, 85, 88–90, 97–9, 103, 105–6, 131, 134–5, 139, 141, 153, 155, 157, 160, 165, 167, 174, 177–82, 186, 188, 190, 197, 199, 201, 207, 216, 219–22, 225, 239, 252–3, 257–8, 260, 265–7, 269–75, 278–9, 280n3, 283–90, 294–8, 304, 314n12, 319–22, 324–5, 328–30
coarse-grained 61
mixed/impure 4, 7, 21, 28, 32n28, 51, 66, 74–5, 78, 90, 260, 262–3, 265, 271–5, 277–9, 286, 295–7 (*see also* mixture; state, pure)
pure 4, 7, 17–18, 27–8, 33n28, 39, 46, 49, 51, 63, 66, 72, 74, 78, 90, 206–8, 260–66, 271–5, 278–9, 280n3, 286–7, 289, 295–7 (*see also* state, mixed/impure)
reduced 46, 60–61, 73, 274
universal/of the universe 6, 10–11, 28–9, 32n24, 39, 48, 51–3, 226, 260–65, 271–9, 289–90, 319, 325–6 (*see also* universe, initial state of)
statistical mechanics/physics 4, 7, 74, 83, 86, 90–91, 94–107, 108n4, 196–7, 330
classical 21, 176, 330

statistical mechanics/physics (*Continued*)
foundations of 89
quantum 75, 80, 82–3, 90, 91n2, 105, 108n6, 272–3, 275, 279
Stern–Gerlach experiment 174, 189, 285, 287–8
Stinespring dilation theorem 31n4, 79, 208–9, 213n10, 297, 299
Stirling's formula 133
structure 9, 22, 25, 28, 47, 49, 59–61, 66, 97–9, 102, 155, 158–9, 161, 163, 169n4, 180, 183, 186, 193–5, 208, 210–11, 216–19, 224, 235, 239, 252–5, 257, 265, 269, 278, 295, 303, 309, 313n1, 314n15, 315n22
causal (of a theory) 2, 98, 319–21
emergent 1, 221
relativistic 97, 103, 319–20
subject–object split 148, 150, 167, 294, 323, 331. *See also* system–observer split
subsystem 4, 7, 10, 18–19, 25, 30, 31n4, 31n7, 51–3, 58–61, 63, 71–5, 78–9, 84, 89–90, 151, 157, 159–60, 163, 165, 233, 237–41, 243, 254, 260–62, 267–75, 279, 280n3. *See also* system
autonomous/independent 6–7, 73, 75, 78–9, 84, 90, 91n2
relativity 7
sun 3, 8, 113–19, 121–4, 125–6n4, 126n14, 203, 218, 318. *See also* solar system; system, earth–moon–sun
superconductivity 80, 82, 97–8
superdeterminism 324–5, 327. *See also* determinism
superfluidity 82, 98
superoscillation 287–8
superposition 5, 27, 32n27, 157, 165, 284, 286–7, 289–90, 321
symmetric criticality, principle of 254
symmetry 40, 62, 66, 79, 82, 108n3, 152–4, 160–61, 167, 169n3, 185, 218–19, 240–41, 252, 254, 284, 288, 320. *See also* asymmetry, temporal/time
dynamical 153
gauge ~ 152–3, 159–60, 164, 166 (*see also* invariance)
principle 240–41
scaling ~ 255
system. *See also* subsystem
adaptive 2, 221
–agent split (*see* system–observer split)
autonomous 7, 10, 39–40, 52, 71–9, 84, 88, 90, 202–3, 205, 207, 209, 250, 253–5, 257
biological 1, 9, 63, 216, 218 (*see also* organism)
closed (*see* closed system; closed systems view; closure)
contact ~ 248, 250–58
earth–moon 203
earth–moon–sun 116–17, 200
isolated 4–5, 7–9, 11, 17–19, 31n6, 39–40, 43, 63, 71–4, 77–8, 83–4, 89–90, 113–16, 121–3, 134, 177, 193–4, 196–201, 203, 205–6, 209–10, 212n4, 245, 283, 294–5
–observer split 148, 150, 167, 317, 323, 331
open (*see* open system; open systems view)
quantum(-mechanical) 4–7, 9, 11, 30n4, 31n13, 43, 46, 51–2, 57–67, 71–91, 94, 96, 137, 156, 165, 174–90, 191n2, 194, 197, 203, 205–10, 212, 212n2, 212n10, 275, 286, 293–7, 299–300, 304–6, 312, 314n16
self-gravitating 81, 247
solar (*see* solar system)

technology 217, 222, 224–5, 226, 285, 309
theory. *See also* framework–theory–model distinction
development 129–30
field ~ 7, 30n4, 48–9, 62, 74, 77, 80, 82, 86, 88, 90, 91n1, 94, 105–6, 107n1, 166, 198, 211, 320, 322, 324–6, 328
fundamental 5, 11, 64, 66, 71, 103–4, 108n4, 114, 127n21, 153, 236–7, 304, 317–19, 324, 331
gauge ~ 153, 155, 159–60, 162–3, 166, 169n3
Type A 95 (*see also* framework)
Type B 95
thermal time hypothesis 265
thermodynamic(s) 1, 4, 106, 130, 196–7, 206, 218, 298, 313n7
limit 83, 293, 298
second law of 136, 193, 216, 220
time 40, 48–9, 52, 57, 73, 88, 98, 100, 108n6, 117, 125n2, 151–4, 165, 169n1, 187, 190, 203, 224–6, 234, 244, 255, 258n4, 265, 289, 296, 300–301, 305, 308–9, 314n12, 314n17, 319, 321–2, 326. *See also* spacetime
arrow of 10, 88–9, 247, 251–2, 258, 258n5, 265
asymmetry (*see* asymmetry, temporal/time)
reversal 88–9, 91n4
slice 6, 38, 41–2, 44–8, 50–53, 256
thermal (*see* thermal time hypothesis)
timescale 10, 75–7, 199–201, 203, 217, 220, 223–6, 303. *See also* scale
TOSV. *See* open systems view, temporally
trace 28, 75–6, 87–8, 206, 208, 275, 296
~ preservation 208, 295

tracing out 4–5, 209, 272–4, 328
trajectory, spacetime 6, 57, 152, 249, 278
transformation, unitary 18, 28
Turing machine 302–3, 307, 315n20
two-chief-world-systems problem 114, 118, 121
typicality 272–4, 280n3

ultraviolet catastrophe 144n4
uncertainty 4, 157, 226n2, 284, 286–7, 289–90, 299
undecidability 303
unit
  fundamental 217, 301
  in a spin network 183–7, 189
unitarity 72–7, 79, 84, 89–90, 187, 189, 206–7, 210, 212n2, 290. *See also* non-unitarity
  puzzle 290
unitary–nonunitary distinction 212n2
universalism, mereological 243
universals, problem of 301, 303
universe 6–7, 9–11, 26–7, 38–41, 44, 46, 49, 51–4, 71, 73, 78, 89–90, 157, 181, 194–5, 197–200, 209–12, 218, 225, 233–4, 237–8, 240–45, 253–4, 256–7, 260–63, 267–8, 272–7, 279, 283–4, 289–90, 295, 300, 305–12, 314n12, 314n17, 315n24, 322, 327
  as a closed system 2, 9, 51–4, 65, 194, 198–9, 209–11, 233, 243, 253, 283, 293, 306, 311
  initial state of 225–6, 258n3, 265, 279, 284, 289–90, 305
  measurement of (*see* measurement of the universe)
  as an open system 9–10, 53, 194, 209–12, 213n11, 217–18, 225–6, 241, 243, 247, 311
  as a plurality 11, 293, 306, 311–12
  quantum state of (*see* state, universal/of the universe)
  wave function of (*see* wave function, universal/of the universe)
unpredictability 216, 219, 221–2, 329

vagueness, fundamental nomic 279
variable, external/exogenous 105–6, 317–18, 321, 328–9 (*see also* parameter, exogenous)
Volta effect 143
von Neumann collapse 283. *See* collapse
von Neumann equation 207, 263, 270, 276. *See also* Liouville–von Neumann equation

wave function 4, 17, 260, 263, 268, 273, 275, 278, 280n3, 283–5, 287–8
  for a cell 226n2
  collapse of (*see* collapse)
  realism (WFR) 11, 47–50, 261–3, 265–8, 272–9 (*see also* density matrix realism)
  universal/of the universe 10–11, 157, 260–79, 283–4
well-posedness 9, 194, 202–7, 212n7
Wentaculus 278–9. *See also* quantum Mentaculus
WFR. *See* wave function realism
Wien's displacement law 134
Wien('s) distribution law 131
Wigner's friend 167, 285
world 10–11, 28, 39, 41–2, 66, 71, 94, 113, 125n1, 236, 260, 265, 293–4, 297–301, 304–5, 308–10, 314n15, 317, 319. *See also* reality; universe
  Everettian 11, 23, 27–9, 283–90
  possible 244, 267 (*see also* possibility)

Yang–Mills theory 154–5, 159–60, 169n3