PHILOSOPHY OF
Quantum Information and Entanglement

EDITED BY

Alisa Bokulich and Gregg Jaeger

CAMBRIDGE www.cambridge.org/9780521898768

This page intentionally left blank

PHILOSOPHY OF QUANTUM INFORMATION AND ENTANGLEMENT

Recent work in quantum information science has produced a revolution in our understanding of quantum entanglement. Scientists now view entanglement as a physical resource with many important applications. These range from quantum computers, which would be able to compute exponentially faster than classical computers, to quantum cryptographic techniques, which could provide unbreakable codes for the transfer of secret information over public channels. These important advances in the study of quantum entanglement and information touch on deep foundational issues in both physics and philosophy.

This interdisciplinary volume brings together fourteen of the world's leading physicists and philosophers of physics to address the most important developments and debates in this exciting area of research. It offers a broad spectrum of approaches to resolving deep foundational challenges – philosophical, mathematical, and physical – raised by quantum information, quantum processing, and entanglement. This book will interest physicists, philosophers of science, and computer scientists.

ALISA BOKULICH is a Professor in the Philosophy Department at Boston University, and the director of Boston University's Center for Philosophy and History of Science. Her research focuses on the history and philosophy of physics, as well as broader issues in the philosophy of science.

GREGG JAEGER is an Associate Professor at Boston University, where he teaches courses in the Mathematics, Natural Science, Philosophy, and Physics departments. His recent research focuses on decoherence, entanglement, quantum computing, and quantum cryptography, and in 2008 he was awarded a Kavli fellowship.

PHILOSOPHY OF QUANTUM INFORMATION AND ENTANGLEMENT

Edited by

ALISA BOKULICH
Boston University

and

GREGG JAEGER
Boston University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521898768

First published in print format 2010

ISBN-13 978-0-511-68024-3 eBook (EBL)

ISBN-13 978-0-521-89876-8 Hardback

For our teacher, mentor, and colleague
Abner Shimony
who has contributed so much
to both the philosophy and the physics of this subject.

Contents

Contributors

Howard Barnum
CCS-3: Modeling, Algorithms, and Informatics,
Mail Stop B256,
Los Alamos National Laboratory,
Los Alamos, NM 87545, USA

Jeffrey Bub
Department of Philosophy,
University of Maryland,
College Park, MD 20742, USA

Giacomo Mauro D'Ariano
Istituto Nazionale di Fisica della Materia, Unità di Pavia,
Dipartimento di Fisica "A. Volta,"
via Bassi 6, I-27100 Pavia, Italy

Armond Duwell
Department of Philosophy,
University of Montana,
Missoula, MT 59812, USA

Lucien Hardy
Perimeter Institute,
31 Caroline Street North,
Waterloo, Ontario N2L 2Y5, Canada

Leah Henderson
Department of Linguistics and Philosophy, MIT,
77 Massachussetts Avenue 32-D808
Cambridge, MA 02139-4307, USA

Andrei Khrennikov
International Center for Mathematical Modeling in Physics and Cognitive Sciences,
University of Växjö, S-35195, Sweden

F. Hadi Madjid
82 Powers Road,
Concord, MA 01742, USA

John M. Myers
Division of Engineering and Applied Sciences,
Harvard University,
60 Oxford Street,
Cambridge, MA 02138, USA

Wayne C. Myrvold
Department of Philosophy,
Talbot College,
University of Western Ontario,
London, Ontario N6A 3K7, Canada

Sandu Popescu
H. H. Wills Physics Laboratory,
University of Bristol,
Tyndall Avenue,
Bristol, BS8 1TL, UK; and
Hewlett-Packard Laboratories,
Stoke Gifford,
Bristol, BS12 6QZ, UK

Christopher G. Timpson
Brasenose College,
University of Oxford, OX1 4AJ, UK

Lorenza Viola
Department of Physics and
Astronomy,
Dartmouth College,
6127 Wilder Laboratory,
Hanover, NH 03755, USA

Tai Tsun Wu
Harvard University,
Cambridge,
MA 02138, USA

Preface

Recently there has emerged an exciting and rapidly growing field of research known as quantum information theory. This interdisciplinary field is unified by the following two goals: first, the possibility of harnessing the principles and laws of quantum mechanics to aid in the acquisition, transmission, and processing of information; and second, the potential that these new technologies have for deepening our understanding of the foundations of quantum mechanics and computation. Many of the new technologies and discoveries emerging from quantum information theory are challenging the adequacy of our old concepts of entanglement, non-locality, and information. This research suggests that the time is ripe for a reconsideration of the foundations – and philosophical implications – of quantum information theory.

Historically, apart from a small group of physicists working on foundational issues, it was philosophers of physics who recognized the importance of the concepts of entanglement and non-locality long before the mainstream physics community. Prior to the 1980s, discussions of the infamous "EPR" paper and John Bell's seminal papers on quantum non-locality were carried out more often by such philosophers than by ordinary physicists. In the 1990s that situation rapidly changed, once the larger community of physicists had begun to realize that entanglement and non-locality were not just quirky features of quantum mechanics, but physical resources that could be harnessed for the performance of various practical tasks. Since then, a large body of literature has emerged in physics, revealing many new dimensions to our concepts of entanglement and non-locality, particularly in relation to information. Regrettably, however, only a few philosophers have followed these more recent developments, and many philosophical discussions still end with Bell's work.

The purpose of this volume is two-fold. First, our hope is to introduce more philosophers of physics to the recent discussions about entanglement and non-locality by making accessible some of the central developments in this field

"beyond Bell." While there are many excellent anthologies examining the philosophical implications of Bell's theorem, the present volume is the first interdisciplinary anthology to explore the philosophical implications of entanglement and non-locality beyond Bell. In this sense the philosophers have much to learn from the physicists. The second goal of this volume is to encourage more physicists to reflect critically on the foundations of quantum information theory. The key concepts of entanglement, non-locality, and information are in need of conceptual clarification and perhaps even bifurcation. Recent claims that quantum information science revolutionizes the foundations of quantum mechanics and solves its most basic conceptual puzzles need to be critically examined. Here the physicists have much to learn from the philosophers, who have long been engaged in such projects of conceptual clarification and logical analysis. Our hope is that pursuing these two goals together will encourage a fruitful dialogue and that a stronger interdisciplinary field in the philosophy of quantum information will be forged.

The idea for this volume first emerged at a conference on the foundations of quantum information and entanglement, which was held at the Center for Philosophy and History of Science at Boston University in 2006. The conference was a tremendous success, drawing over two hundred attendees, and it emphasized the need for an interdisciplinary volume in this area. Many of the speakers at this conference were chosen to be contributors to the present volume. We gratefully acknowledge the support of the Center for Philosophy and History of Science and the National Science Foundation for making this conference possible. We would also like to thank Molly Pinter for her work compiling the index.

We have gathered here twelve original papers, seven of which are by physicists and five of which are by philosophers, all of whom are actively engaged in quantum information theory. These papers, which are by many of the leading researchers in the field, represent a broad spectrum of approaches to the foundations of quantum information theory and highlight some of the most important developments and debates. While these papers assume a certain level of scientific literacy, an effort has been made to present the latest research in a way that is accessible to non-specialists, physicists and philosophers of physics alike. To this end, the volume begins with a pedagogical introduction, briefly laying out the relevant historical background, as well as defining the key philosophical and physical concepts used in the subsequent papers. While it would be impossible to cover all the important developments in this rapidly growing field, we hope this volume succeeds in laying the foundation for further interdisciplinary work in the philosophy of quantum information and entanglement, by encouraging more physicists and philosophers to enter into the debate.

Introduction

Entanglement can be understood as an extraordinary degree of correlation between states of quantum systems – a correlation that cannot be given an explanation in terms of something like a common cause. Entanglement can occur between two or more quantum systems, and the most interesting case is when these correlations occur between systems that are space-like separated, meaning that changes made to one system are immediately correlated with changes in a distant system even though there is no time for a signal to travel between them.[1] In this case one says that quantum entanglement leads to non-local correlations, or non-locality.

More precisely, entanglement can be defined in the following way. Consider two particles, A and B, whose (pure) states can be represented by the state vectors ψ_A and ψ_B. Instead of representing the state of each particle individually, one can represent the composite two-particle system by another wavefunction, Ψ_{AB}. If the two particles are *unentangled*, then the composite state is just the tensor product of the states of the components: $\Psi_{AB} = \psi_A \otimes \psi_B$; the state is then said to be factorable (or separable). If the particles are entangled, however, then the state of the composite system *cannot* be written as such a product of a definite state for A and a definite state for B. This is how an entangled state is defined for pure states: a state is entangled if and only if it cannot be factored: $\Psi_{AB} \neq \psi_A \otimes \psi_B$. For mixed states, which must be represented by density operators rather than state vectors, the definition of entanglement is generalized: an entangled mixed state is one that *cannot* be written as a convex combination of products

$$\rho_{AB} = \sum_i p_i (\rho_{Ai} \otimes \rho_{Bi}),$$

[1] On some conceptions, entanglement can occur even between the different properties of a single quantum system, such as in the case of entangling a particle's position with its spin.

Philosophy of Quantum Information and Entanglement, ed. A. Bokulich and G. Jaeger. Published by Cambridge University Press.

where the sum of the p_i is equal to unity. This definition is for a bipartite system, that is, a composite system of only two parts, A and B. For multipartite mixed quantum systems the situation is more complicated; there is no single acceptable entanglement measure applicable to the full set of possible states of systems having a greater number of parts.[2] The search for a fully general definition and measure of entanglement remains an active area of research.

Despite the fact that the phenomenon of entanglement was recognized very early on in the development of quantum mechanics, it remains one of the least understood aspects of quantum theory. It was arguably the well-known "EPR" paper,[3] by Albert Einstein, Boris Podolsky, and Nathan Rosen, published in May of 1935, that first drew attention to the phenomenon of entanglement.[4] For EPR, however, the possibility of such a phenomenon in quantum mechanics was taken to be a *reductio ad absurdum* showing that there is a fundamental flaw with the theory: "since at the time of measurement the two systems no longer interact, no real change can take place in the second system in consequence of anything that may be done to the first system" (Einstein *et al.* 1935, p. 779); since quantum mechanics implies such an "absurd" situation, quantum mechanics must be incomplete at best. Quantum entanglement, however, precisely is such a non-classical relationship between quantum particles whereby changes made to one particle of an entangled pair can lead to changes in the other particle even though they no longer interact.

Shortly after the appearance of the EPR paper, Erwin Schrödinger coined the term "entanglement" (*Verschränkung*) to describe this phenomenon. The first published occurrence of the term is in an article of his, written in English, which appeared in October of 1935. In this article, Schrödinger places the phenomenon of entanglement at the center of quantum theory:

> When two systems, of which we know the states by their respective representatives, enter into temporary physical interaction due to known forces between them, and when after a time of mutual influence the systems separate again, then they can no longer be described in the same way as before, viz. by endowing each of them with a representative of its own. I would not call that *one* but rather *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought. By the interaction the two representatives (or ψ-functions) have become entangled (Schrödinger 1935a, p. 555).

The term appears two months later in German in the second of the triplet of papers in which he introduces his infamous cat paradox (Schrödinger 1935b).

[2] For a technical overview of basic results in entanglement and quantum information see, for example, Jaeger (2007).

[3] See Fine (2008) for a helpful overview of the EPR paper.

[4] Regarding the question of when the physics community first became aware of the phenomenon of entanglement, Don Howard (1990) has cogently argued that Einstein had recognized, and been concerned about, the phenomenon of entanglement long before the 1935 EPR paper.

Although Schrödinger recognized the importance of the phenomenon of entanglement, he certainly did not embrace this feature of quantum theory. His dissatisfaction with the phenomenon of entanglement, and with the quantum theory in general, is evident in his correspondence with Einstein during this time. In a letter to Einstein, Schrödinger writes as follows regarding the recently published EPR paper: "I was very happy that, in your work that recently appeared in *Phys. Rev*., you have publicly caught the dogmatic quantum mechanics by the collar, regarding that which we had already discussed so much in Berlin" (Schrödinger to Einstein, June 7, 1935).[5] Toward the end of the same letter he continues, "The point of my foregoing discussion is this: we do not have a quantum mechanics that takes into account relativity theory, that is, among other things, that respects the finite speed of propagation of all effects" (Schrödinger to Einstein, June 7, 1935). Schrödinger's concern is that the phenomenon of entanglement, exhibiting as it does non-local correlations between separated particles, ultimately would prove to be in conflict with the first-signal principle of special relativity.

Despite this early recognition of the importance of the phenomenon, very little effort or progress was made over the next thirty years in developing a theory of entanglement or in answering Schrödinger's concerns regarding how this phenomenon could be consistent with relativity. It would be almost thirty years before another significant step toward a theory of entanglement would be made with John Bell's seminal 1964 paper on quantum non-locality. In that paper Bell considered a pair of particles in the singlet state that had interacted in the past, had become entangled, and then had separated. He derived an inequality involving the probabilities of various outcomes of measurements performed on these entangled particles that any local definite (i.e., hidden-variable) theory must satisfy. He then showed that quantum mechanics violates this inequality; that is, the experimentally well-confirmed quantum correlations among entangled particles cannot be locally explained. Bell's theorem does not rule out the possibility of hidden-variable theories in general, only those hidden-variable theories that are local. Indeed, Bell took the lesson of his theorem to be that any theory that reproduces the experimentally well-confirmed predictions of quantum mechanics must be non-local. He writes,

> It is the requirement of locality, or more precisely that the result of a measurement on one system be unaffected by operations on a distant system with which it has interacted in the past, that creates the essential difficulty . . . This [non-locality] is characteristic, according to the result to be proved here, of any theory which reproduces exactly the quantum mechanical predictions (Bell 1964, p. 14).

[5] This translation is taken from a copy of the letter located at the Howard Gotlieb Archival Research Center at Boston University. The original letter is in German and is held in the Einstein Archives at the Hebrew University in Jerusalem.

What is remarkable about Bell's theorem is that it is a general result arising from an analysis of the relevant probabilities of various joint measurement outcomes, and does not depend on the details of any hidden-variable theory or even on the details of quantum mechanics itself.[6] Since then a number of different Bell-type inequalities have been derived, such as the Clauser, Horne, Shimony, and Holt (CHSH) inequality (1969), which has proven particularly useful for experimental tests of non-locality. Following Bell, a number of experiments demonstrated not only that non-locality is a genuine physical phenomenon characteristic of our world (e.g., Aspect *et al.* 1982), but also that non-locality can be experimentally produced, controlled, and harnessed for various applications.

Another theoretical development came with Jon Jarrett's (1984) analysis showing that Bell's locality condition can be viewed as the conjunction of two logically independent conditions: a "controllable" locality, which if violated would conflict with special relativity, and an "uncontrollable" locality whose violation might "peacefully coexist" with relativity (see also Shimony (1984); for an opposing point of view see Maudlin (1994)). Hence, the violation of Bell's inequality could logically be due to a violation of one, the other, or both of these locality conditions. Jarrett's analysis has been taken by some to provide the solution to Schrödinger's worries about a conflict between quantum theory and relativity, as long as one assumes that the violation is in fact solely a violation of the uncontrollable locality.[7]

Despite these important advances, it was still only a handful of physicists who were deeply interested in entanglement. Philosophers of physics recognized the importance of entanglement and Bell's work, but many continued to think of entanglement as an "all or nothing" phenomenon and described entanglement as simply a spooky action-at-a-distance or mysterious holism. Moreover, the bulk of philosophical work on non-locality and entanglement has considered developments only up to and including Jarrett's analysis and the experiments performed in the mid 1980s. In the last two decades new discoveries – many of which are associated with the investigation of quantum information – have shown that much philosophical and foundational work remains to be done to deepen our understanding of entanglement and non-locality.

Toward the end of the 1980s and the beginning of the 1990s a number of important transformations in our understanding of entanglement took place. First, it was recognized (e.g., Shimony (1995)) that entanglement can be quantified; that is,

[6] There are of course many subtleties in analyzing the implications of Bell's theorem that cannot be covered in this introduction, but are discussed extensively in the voluminous literature that followed Bell's work. For an introduction to the subtleties of interpreting the lessons of Bell's theorem see Cushing and McMullin (1989) and Shimony (2008).

[7] For an overview of these issues see Cushing and McMullin (1989).

it comes in degrees ranging from "maximally entangled" to not entangled at all. Moreover, entanglement can be manipulated in all sorts of interesting ways. For example, Bennett *et al.* (1996) have shown that one can take a large number of electrons that are all partly (that is, "a little bit") entangled with each other, and concentrate that entanglement into a smaller number of maximally entangled electrons, leaving the other electrons unentangled (a process known as *entanglement distillation*).[8] Conversely, one can take a pair of maximally entangled electrons and spread that entanglement out over a larger number of electrons (so that they are now only partly entangled) in such a way that the total entanglement is conserved (a process known as *entanglement dilution*).

The notion of a "degree of entanglement" seems to have been first recognized through the related notion of a degree of violation of the Bell inequalities – indeed, this was used as the first measure of entanglement in the case of pure states: the greater the degree of violation of the inequalities, the greater the amount of entanglement. Nicolas Gisin describes this "quiet revolution" as follows:

> In this brief note I prove that the product states are the only states that do not violate any Bell inequality. When I had the chance to discuss this equivalence between "states that violate the inequality" and "entangled states" (i.e., "non-product states") with John Bell last September, just before his sudden tragic death, I was surprised that he did not know this result. This motivates me to present today this little note which I have had on my shelves for many years and which may be part of the "folklore," known to many people but (apparently) never published. (Gisin 1991, p. 201)

There are, however, limitations to using a violation of Bell's inequality as a general measure of entanglement. First, there are Bell-type inequalities whose largest violation is given by a non-maximally entangled state (Acín *et al.* 2002), so entanglement and non-locality do not always vary monotonically. More troublingly, however, Reinhard Werner (1989) showed that there are some mixed states (now referred to as Werner states) that, though entangled, do not violate Bell's inequality at all; that is, there can be entanglement without non-locality. In an interesting twist, Sandu Popescu (1995) has shown that even with these local Werner states one can perform a non-ideal measurement (or series of ideal measurements) that "distills" a non-local entanglement from the initially local state. In yet a further twist, the Horodecki family (1998) subsequently showed that not all entanglement can be distilled in this way – there are some entangled states that are "bound." These bound entangled states are ones that satisfy the Bell inequalities (i.e., they are local) and cannot have maximally entangled states violating Bell's inequalities extracted from them by means of local operations.[9]

[8] It is also sometimes referred to as "entanglement concentration" or "entanglement purification."
[9] For a nice review of these developments see Werner and Wolf (2001).

Not only can one have entanglement without non-locality, but also, as Bennett *et al.* (1999) have shown, one can have a kind of "non-locality without entanglement." There are systems that exhibit a type of non-local behavior even though entanglement is used neither in the preparation of the states nor in the joint measurement that discriminates the states (see also Niset and Cerf (2006)). This work highlights another facet of the concept of non-locality, which, rather than involving correlations for space-like separated systems, involves instead a kind of indistinguishability based on local operations and classical communication. The relationship between this new notion of non-locality and the traditional one involving space-like separated systems remains to be worked out.

These recent developments point to the need for a new, more adequate way of measuring and quantifying entanglement. They show that the concepts of entanglement and non-locality are much more subtle and multifaceted than earlier analyses based solely on Bell's theorem realized. Much philosophical and foundational work remains to be done on understanding precisely how the important notions of entanglement and non-locality are related.

These questions of how to quantify entanglement and non-locality – and the need to clarify the relationship between them – are important not only conceptually, but also practically, insofar as entanglement and non-locality seem to be different resources for the performance of quantum information processing tasks. As Brunner and colleagues have argued, it is important to ask "whether in a given quantum information protocol (cryptography, teleportation, and algorithm . . .) it is better to look for the largest amount of entanglement or the largest amount of non-locality" (Brunner *et al.* 2005, p. 12). Arguably it is this new emphasis on the exploitation of entanglement and non-locality for the performance of practical tasks that marks the most fundamental transformation in our understanding of these concepts.

The newly formed field of quantum information theory is devoted to using the principles and laws of quantum mechanics to aid in the acquisition, transmission, and processing of information. In particular, it seeks to harness the peculiarly quantum phenomena of entanglement, superposition, and non-locality to perform all sorts of novel tasks, such as enabling computations that operate exponentially faster or more efficiently than their classical counterparts (via quantum computers) and providing unconditionally secure cryptographic systems for the transfer of secret messages over public channels (via quantum key distribution). By contrast, classical information theory is concerned with the storage and transfer of information in classical systems. It uses the "bit" as the fundamental unit of information, where the system capable of representing a bit can take on one of two values (typically 0 or 1). Classical information theory is based largely on the concept of information formalized by Claude Shannon in the late 1940s. Quantum information theory, which was later developed in analogy with classical information theory, is concerned with the

storage and processing of information in quantum systems, such as the photon, electron, quantum dot, or atom. Instead of using the bit, however, it defines the fundamental unit of quantum information as the "qubit." What makes the qubit different from a classical bit is that the smallest system capable of storing a qubit, the two-level quantum system, not only can take on the two distinct values $|0\rangle$ and $|1\rangle$, but can also be in a state of superposition of these two states: $|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle$.

Quantum information theory has opened up a whole new range of philosophical and foundational questions. The first cluster of questions concerns the nature of quantum information. For example, is quantum information just classical information stored in a quantum system, or is it a new distinctive type of information? (See Chapter 8 by Duwell.)

A second cluster of important philosophical questions concerns how it is that quantum information protocols are able to achieve more than their classical counterparts. For example, how is that quantum computers are able to compute exponentially faster than classical computers? (See Chapter 11 by Bub.) Can quantum computers perform calculations that are not Turing computable – so-called hypercomputation? Another example concerns quantum teleportation, whereby the complete state of a quantum system (something that would take a huge amount of information to specify classically) can be transferred to another distant system using only two bits of information, as long as the two parties at the different locations share a pair of entangled particles. The *prima facie* puzzle of teleportation is how so much "information" can be transferred so efficiently (see Penrose (1998), Deutsch and Hayden (2000), and, for critical analyses, Timpson (2006) and Jaeger (2009)).

Yet another example of a quantum information technology raising foundational questions is quantum cryptography or quantum key distribution, which involves using the principles of quantum mechanics to ensure secure communication (that is, the transfer of secret information over public channels in a way that cannot be successfully eavesdropped upon). Some quantum cryptographic protocols make use of entanglement to establish correlations between systems that would be lost upon eavesdropping. Moreover, a quantum principle known as the no-cloning theorem prohibits making identical copies of an unknown quantum state.[10] This theorem ensures that an eavesdropper cannot make a copy of the cryptographic key without the communicating parties knowing that this is happening. One important question is whether these quantum principles are really sufficient to provide unconditional security, that is security in the face of all conceivable attacks. (See Chapter 6 by Myers and Madjid, and Chapter 12 by Wu.)

[10] Quantum teleportation, mentioned earlier, is not in conflict with the no-cloning theorem since the initial state is automatically destroyed upon teleportation, that is, it does not involve cloning an unknown quantum state.

A third important cluster of philosophical questions concerns what new insights recent work in quantum information theory might provide into the foundations of quantum mechanics. Some authors have argued that an information-theoretic approach may provide a new axiomatic basis for quantum mechanics and provide deeper insight into what makes quantum mechanics different from classical mechanics. Anton Zeilinger (1999) has proposed a new information-theoretic "foundational principle," which he believes can explain both the intrinsic randomness of quantum theory and the phenomenon of entanglement. (For critical discussions see Chapter 10 by Timpson and Jaeger (2009).) In another approach, Chris Fuchs (2002) has adopted a Bayesian approach and argued that quantum mechanics just is quantum information theory – a more sophisticated gloss on the old idea that a quantum state is just a catalogue of expectations. (For a discussion of the Bayesian approach see Chapter 7 by Henderson.) Yet a third approach that has generated considerable discussion is a theorem proven by Rob Clifton, Jeff Bub, and Hans Halvorson (2003). In the context of a C^*-algebraic formulation, they argue that quantum theory can be characterized in terms of three information-theoretic constraints: (1) no superluminal signaling via measurement, (2) no cloning (for pure states) or no broadcasting (mixed states), and (3) no unconditionally secure bit commitment.[11] (For a discussion of the relative strengths and weaknesses of this approach see Chapter 9 by Myrvold.) Bub (2004) in particular has taken this ("CBH") theorem to show that quantum theory is best interpreted as a theory about the possibilities of information transfer rather than a theory about the non-classical mechanics of waves or particles. Much philosophical work remains to be done assessing these various claims that quantum information provides a new, more adequate way of conceiving quantum theory.

All the contributors to this volume are grappling with different facets of the challenges and opportunities that quantum information theory poses for quantum mechanics. The papers are organized into the following four topics: quantum entanglement and non-locality, quantum probability, quantum information, and quantum communication and computing.

The three papers in the first section are concerned with expanding and generalizing the central notions of entanglement and non-locality. The first paper, by Sandu Popescu, explores the notion of *degrees of non-locality* by considering the possibility of theories that exhibit even stronger non-local correlations than quantum

[11] Bit commitment is a central cryptographic protocol between two mistrusting parties (typically referred to as Bob and Alice) in which Bob obtains an encoded bit from Alice. It is secure against Bob if Bob cannot decode the bit until Alice chooses to reveal it by supplying some further information, and it is secure against Alice if Alice cannot change the bit (it is fixed between commitment and revealment). A theorem by Mayers (1997) showed that there is no unconditionally secure (i.e., secure against any conceivable attack) standard quantum bit commitment.

mechanics and yet still don't violate the first-signal principle of relativity. One can see the possibility of these "super-quantum" correlations through conceivable degrees of violation of the Clauser–Horne–Shimony–Holt (CHSH) inequality (Clauser *et al.* 1969). Like Bell's original inequality, the CHSH inequality sets a bound on a particular linear combination of the set of correlations, $E(X, Y)$, between two parties, A and B, that must be respected by any (Bell-)local theory:

$$-2 \leq E(A, B) + E(A, B') + E(A', B) - E(A', B') \leq 2.$$

If a theory produces correlations whose sum exceeds the upper bound of 2 then the theory is said to be (Bell) non-local. A theorem by Tsirel'son (also sometimes transliterated Cirel'son) states that the maximum bound on the correlations quantum mechanically is $2\sqrt{2}$; however, assuming only no signaling, the upper bound on the correlations could mathematically be as high as 4, and it is the region in between these two bounds that defines the superquantum correlations, or non-local correlations that are stronger than quantum mechanics.

A useful tool for investigating degrees of non-locality is the "PR box" named after Sandu Popescu and Daniel Rohrlich, who first formalized it in response to a question posed by Abner Shimony concerning whether the conjunction of the conditions of causality and no signaling uniquely picks out quantum mechanics from all possible correlation-predicting theories. The PR box can be thought of as a "black-box" device to which each of the two parties, A and B, has access to half of. A and B can each select an input from a range of possibilities and then obtain a particular output (which they cannot control) from the box. The central function of the box is to determine the joint probability of the two outputs given the two inputs.

An experimental arrangement to measure a quantum system in a particular state can be thought of as one example of such a box, where the input is a particular measurement choice at each wing A and B (such as measuring the polarization of a photon along a certain direction) and the output is a certain measurement outcome (such as getting a horizontal polarization). The no-signaling requirement is imposed on the boxes by requiring that the outcome at A is independent of the measurement choice at B. In their 1994 paper, Popescu and Rohrlich wrote down a correlation function for a set of measurements that yielded a value of 4 for the left-hand side of the CHSH inequality and yet still prohibited signaling, suggesting that quantum mechanics was just one among a set of possible non-local theories that are consistent with relativity theory.

These super-quantum correlations are particularly interesting from an information-theoretic point of view insofar as they would radically reduce the amount of communication needed for distributed computational tasks (Barrett *et al.* 2005). In his paper for this volume (Chapter 1), Popescu further explores what

advantage superquantum correlations would provide for performing various quantum communication and computing tasks. In particular, Popescu examines "non-local computation" whereby separated parties A and B must compute a function without either party learning anything about the inputs. Popescu demonstrates that, while neither classical mechanics nor quantum mechanics permit such non-local computations to succeed, any theory with non-locality even just infinitesimally stronger than quantum mechanics does allow non-local computation to take place.

The second contribution to this volume focuses on the concept of entanglement and how the notion of entanglement might be generalized for situations in which the overall system cannot be easily partitioned into separated subsystems A and B. The standard definition of entanglement for pure states depends on being able to define two or more subsystems for which the state cannot be factored into product states. For strongly interacting quantum systems, such as indistinguishable particles (bosons or fermions) that are close enough together for quantum statistics to be important, the entangled systems cannot easily be partitioned into subsystems in this way. In response to this problem, Lorenza Viola and Howard Barnum have developed a notion of "generalized entanglement," which depends on the expectation values of a preferred set of observables, rather than on a partitioning of the entangled system into subsystems. The intuition behind their approach is that entangled pure states look mixed to local observers, and the corresponding reduced state provides expectation values for a set of distinguished observables. They define a pure state as "generalized unentangled" relative to the distinguished observables if the reduced state is pure and "generalized entangled" otherwise (Barnum *et al.* 2004, p. 1). Similarly a mixed state is "generalized unentangled" if it can be written as a convex combination of unentangled pure states. Their hope is that this new approach will lead to a deeper understanding of entanglement by allowing it to be defined in more general contexts.

In the third chapter of this volume, Lucien Hardy explores how the concepts of entanglement and information flow will likely have to change in light of attempts to develop a quantum theory of gravity. Quantum mechanics and general relativity – though two of our most successful and well-confirmed scientific theories – are currently inconsistent with one another in certain respects: general relativity is deterministic but has a non-fixed causal structure, while quantum mechanics is inherently indeterministic but has a fixed causal structure. The hope is to find a quantum theory of gravity that unifies these two theories as limiting cases, and Hardy's bet is that such a theory will be indeterministic (probabilistic) and yet have an indefinite causal structure. In a theory with indefinite causal structure, there will be no fact of the matter about whether two systems are space-like separated, for example. Hence the notion of entanglement, which requires two space-like separated systems, and the notion of information flow, which requires a sequence of

time-like related regions, will have to be radically modified. Hardy develops a new formulation of quantum mechanics in terms of what he calls the causaloid framework and then shows how entanglement and information flow can be redefined. He demonstrates how the quantum theory of pairwise interacting qubits can be formulated in the causaloid framework, permitting universal quantum computation.

The second section of this book, on "quantum probability," contains four chapters examining various aspects of the central role that probability theory plays in quantum theory and quantum information science. Not only is quantum mechanics a probabilistic theory, but also the probabilities occurring in quantum mechanics are non-standard probabilities, whose conceptual basis has been an ongoing source of controversy ever since the theory's introduction. Moreover, in the more recent context of quantum information theory, the entropy functions involved in quantifying information in the classical and quantum contexts derive from different sorts of probability, which have distributions of different mathematical forms. Hence, analyses of probability are playing a central role in reexaminations of the foundations of quantum mechanics and quantum information theory.

In Chapter 4, Andrei Khrennikov argues that the challenges currently facing quantum information science point to the need for a reconsideration of the very foundations of quantum mechanics. For example, the security of quantum cryptographic protocols depends on the assumption that standard quantum mechanics is complete and that the quantum probabilities involve irreducible randomness. Khrennikov argues that what is required for quantum information science to move forward is a more rigorous mathematical formulation of probability theory. Khrennikov adopts the controversial view that the experimental violations of Bell-type inequalities do not in fact demonstrate quantum non-locality because the probabilities involved in measurements to test the inequalities are not mathematically well defined. After providing a more rigorous mathematical formulation of quantum probability, he concludes that the lesson of Bell-type "no-go" theorems needs to be modified.

Recent developments in quantum information theory have renewed interest in finding a new axiomatic formulation of quantum mechanics. In his paper for this volume, Giacomo Mauro D'Ariano takes up this challenge of finding a new axiomatization. He begins by noting some of the shortcomings of other recent axiomatizations such as Hardy's (2001) and the much discussed Clifton, Bub, and Halvorson (CBH) derivation of quantum mechanics from three information-theoretic constraints (Clifton *et al.* 2003). D'Ariano argues that a more promising approach to an operational axiomatization involves situating quantum mechanics within the broader context of probabilistic theories whose non-local correlations are stronger than quantum mechanics and yet are still non-signaling (see Chapter 1 by Popescu). He outlines such an axiomatization

in which quantum mechanics is understood as the mathematical representation of a set of rules enabling experimenters to make predictions regarding future events on the basis of suitable tests – an approach he calls the "fair operational framework."

In Chapter 6, John Myers and Hadi Madjid begin by exploring the relation between quantum-mechanical operators and the outcome probabilities these operators generate. In any quantum experiment there is a state preparation, described by a density operator, and a measurement, described by a set of detection operators. Both these operators depend on parameters, which represent the choices made by the experimenters. The trace rule can then be used to determine which parameterized probabilities are the result of a given parameterized density operator and a given parameterized detection operator. After reviewing their recent result proving that any given parameterized probability can be generated by infinitely many different parameterized operators, Myers and Madjid are led to consider parameterized probability measures independently, as a useful object of study in their own right. In their contribution to this volume, Myers and Madjid show how a consideration of these parameterized probability measures leads to three important results for quantum information theory. First, they are able to strengthen Holevo's bound on quantum communication. Holevo's bound is a theorem proving that, even though an arbitrarily large amount of classical information can be encoded in a "qubit," (more precisely the state of a quantum two-level system), such as in the process of defining the direction of a quantum state vector, at most one classical bit of information can be accessed through a measurement of that state (more precisely the accessible information is limited by the von Neumann entropy). Myers and Madjid are able to strengthen the Holevo bound by deriving a stronger inequality limiting the accessible information even in cases for which the von Neumann entropy is arbitrarily large (or infinite), making the traditional formulation of the bound uninformative. Second, they show how this approach can reveal vulnerabilities in quantum key-distribution protocols. Finally, they show that an examination of the parameterized probability measures generated by entangled states can reveal hitherto overlooked topological features, thus deepening our understanding of entangled states.

Another way in which considerations of probability have been at the center of foundational debates in quantum information theory is in the analogy that has been drawn between Bayesian conditionalization and quantum state updating upon measurement (e.g., Bub (1977) and Fuchs (2002)). In the Bayesian approach, named for the eighteenth-century mathematician and theologian Thomas Bayes, probabilities are interpreted as subjective degrees of belief, rather than frequencies. According to Bayes' theorem, or rule, the probability of a hypothesis, H, given some new data, D, is equal to the probability of that data given the hypothesis (i.e., the

conditional probability or "likelihood") times the prior probability of the hypothesis (the "prior"), all divided by the marginal probability of the data:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}.$$

In other words, Bayes' rule tells you how to go about updating a probability distribution in light of new evidence. *Prima facie* there is an analogy between Bayesian updating and quantum measurements insofar as the quantum state gives a set of probabilities for various possible measurement outcomes, and once a measurement is performed information is gained and the probabilities are updated. In Chapter 7 of this volume Leah Henderson offers a critical analysis of this analogy. Drawing on the observation that an efficient quantum measurement is not just a refinement of the probability distribution but also involves an extra unitary transformation, she argues that there is an important disanalogy. Henderson proves that the measurements which can be interpreted as a Bayesian refinement plus a unitary transformation are precisely those measurements which increase our information about the quantum state, and conversely those measurements which do not fall into this category are quantum measurements in which information is actually lost. Such measurements, which increase the uncertainty about the state of the quantum system being measured, are shown to have no direct classical analogue.

The third section of this book turns from foundational questions about probability to foundational questions about the notion of information. In Chapter 8, Armond Duwell tackles head on the question of what precisely quantum information is. There has been considerable debate in the literature over whether quantum information just is classical information stored in quantum systems or whether the classical notion of information, as elaborated by Claude Shannon (1948), is somehow inadequate in this new context. If the classical conception is inadequate, then the question becomes that of what new notion of information should replace it? In his contribution to this volume, Duwell defends what is known as the Schumacher concept of quantum information, following the coding theorem of Ben Schumacher (1995). Duwell divides this notion of quantum information into two parts: quantum quantity-information, which quantifies the resources required to communicate, and quantum type-information, which is the kind of token required to be reproduced at the destination of a communication according to the success criterion of entanglement fidelity (see Duwell (2008) for further details). After discussing the theorem of Clifton, Bub, and Halvorson (2003), which derives quantum theory from three information-theoretic constraints in the context of a C^*-algebraic framework, Duwell criticizes a proposal by Bub that quantum mechanics should be reconceived as a theory of quantum information. Specifically, he argues that Bub

fails to define what notion of quantum information he is using. In a move sympathetic to Bub's approach, Duwell substitutes his own Schumacher notion of quantum information into Bub's proposal and explores the advantages of reconceiving quantum theory in this way.

Quantum information theory is concerned with exploiting the peculiarly quantum features of quantum mechanics to store, process, and transmit information in ways that cannot be achieved classically. This raises the important perennial question of precisely what features of quantum mechanics distinguish it from classical mechanics. Indeed, recent work in quantum information theory has revealed that many features that were thought to be peculiarly quantum turn out to have a classical analogue. In Chapter 9, Wayne Myrvold takes up the task of discovering what it is that makes quantum mechanics distinctive. In his search for the differences, he considers two neutral frameworks in which the classical and quantum theories can be formulated: the algebraic approach and the convex-set approach. He considers a toy theory developed by Rob Spekkens, which he argues reveals some of the key differences between these theories when considered in the context of the convex-set approach. Myrvold draws the intriguing conclusion that, while Schrödinger was right to identify the treatment of compound systems as the distinguishing feature of quantum mechanics, he was wrong to identify entanglement per se as what is distinctively quantum.

It has been argued that quantum information theory may hold the key to solving the conceptual puzzles of quantum mechanics. In Chapter 10, Chris Timpson takes stock of such proposals, arguing that many are just the old interpretative positions of immaterialism and instrumentalism in new guise. Immaterialism is the philosophical view that the world at bottom consists not of physical objects but of immaterial ones – in this context, the immaterial stuff of the world is information. As Timpson shows, this immaterialist view can be seen underlying John Wheeler's (1990) "It from bit" proposal and Zeilinger's "foundational principle" (1999). Similarly, instrumentalism is another philosophical approach that it has long been popular to invoke in the context of quantum mechanics, and has found new life in the context of quantum information theory. Instrumentalism is the view that the task of scientific theories is simply to provide a tool for making predictions – not to be a description of the fundamental objects and laws actually operating in the world. In this context instrumentalism argues that the quantum state is merely a representation of our information, one that allows us to make predictions about experiments, but which should not be thought of as a description of any objective features of the world. Timpson argues that merely re-dressing these well-worn philosophical positions in the new language of information theory does not in fact gain any interpretive ground. After providing a detailed critical analysis of Zeilinger's foundational approach, Timpson concludes that there is indeed

great promise for gaining new insights into the structure and axiomatics of quantum mechanics by focusing on information-theoretic phenomena, as long as one steers clear of the non-starters of immaterialism and instrumentalism.

The final section of the book, on "quantum communication and computing," examines some of the philosophical and foundational questions arising from the new technologies that are emerging from quantum information theory. One of the most tantalizing technologies promised by quantum information theory is the quantum computer. A quantum computer is a computer that exploits the peculiarly quantum features of quantum systems to aid in the processing of data. Much of the interest in quantum computing arose when Peter Shor (1994) devised an algorithm showing that a quantum computer could in principle factor large numbers into primes exponentially faster than any conceivable classical computer. This application is particularly interesting because many current cryptographic protocols for keeping information secure depend on the fact that classical algorithms for factoring take exponentially long; hence, if such a quantum computer were realized, it could pose a threat to the security of the large quantities of information protected in this way.

A few other quantum algorithms have been devised for performing various computations in ways superior to their classical counterparts. Although there are practical issues surrounding the implementation of a quantum computer, one of the key foundational questions is that of determining which feature of quantum mechanics is responsible for the superior computing power of quantum computers. Surprisingly, there is very little agreement over how to answer this question: some have claimed that the speed-up is due to the superposition rule, some attribute it to entanglement, and yet others have claimed that the speed-up of a quantum computer is direct evidence for the so-called "many-worlds" interpretation of quantum mechanics (Deutsch (1997); for critical reviews see Duwell (2007) and Jaeger (2009)). In Chapter 11, Jeffrey Bub proposes a new answer to the question of where the speed-up comes from. According to Bub, the key lies in the difference between classical logic and quantum logic. More specifically, while a classical disjunction is true (or false) by virtue of the truth values of its disjuncts, a quantum disjunction can be true (or false) without any of its disjuncts taking on a truth value at all. Similarly, in the quantum case, a global or disjunctive property of a function is encoded as a subspace of Hilbert space, and a quantum state can end up in a particular subspace without representing any particular pair of input–output values. This is in contrast to a classical computation, in which a global property is represented as a subset, and a classical state can end up in that subset only by ending up at a particular point in the subset (which requires a lot more information, to exclude other points in the subset). He argues that it is not that the quantum algorithm is somehow computing

all values of a function at once that makes it more efficient, but rather that it is in a sense able to avoid computing any values of the function at all.

In the final contribution to the volume, Tai Tsun Wu argues that we need to fundamentally rethink the way we model quantum computing and quantum cryptography. In particular, he argues that the notion of a quantum memory (or quantum register) needs to be included. The content of a quantum memory is a pure state that gets updated to another pure state during a computation via a unitary transformation. The most natural way to model this updating is as a scattering interaction, which is described by the Schrödinger equation and takes the spatial variable explicitly into account. Wu argues that this more physically realistic way of modeling quantum memory leads to a number of surprising results. For example, in the case of quantum key distribution, an analysis of quantum memory using scattering reveals new insecurities. Through a careful examination of the "B92" protocol of Bennett, Wu shows that, by using scattering with one or more spatial variables, forbidden operations such as quantum cloning actually become possible.

As we have seen in this brief overview, quantum information science is in the process of transforming our understanding of both quantum mechanics and information theory. The papers collected in this volume mark an important first step, though there remain many more questions to be explored. Our hope is that this volume will provide a useful starting point for those entering this new interdisciplinary field, and will encourage more philosophers and physicists to enter into the dialogue on the exciting philosophical implications of quantum information research.

References

Acín, A., T. Durt, N. Gisin, and J. Latorre (2002), Quantum non-locality in two three-level systems, *Phys. Rev. A* **65**, 052325.

Aspect, A., J. Dalibard, and G. Roger (1982), Experimental test of Bell's inequalities using time-varying analyzers, *Phys. Rev. Lett.*, **49**, 1804–1807.

Barnum, H., E. Knill, G. Ortiz, R. Somma, and L. Viola (2004), A subsystem-independent generalization of entanglement, *Phys. Rev. Lett.* **92**, 107902.

Barrett, J., N. Linden, S. Massar, S. Pirionio, S. Popescu and D. Roberts (2005), Nonlocal correlations as an information-theoretic resource, *Phys. Rev.* A **71**, 022101.

Bell, J. S. (1964), On the Einstein–Podolsky–Rosen paradox, *Physics* **1**, 195–200. (Reprinted in J. S. Bell (1993), *Speakable and Unspeakable in Quantum Mechanics* (Cambridge: Cambridge University Press), pp. 14–21.)

Bennett, C., G. Brassard, S. Popescu, B. Schumacher, J. Smolin, and W. Wootters, (1996), Purification of noisy entanglement and faithful teleportation via noisy channels, *Phys. Rev. Lett.* **76**, 722–725.

Bennett, C., D. DiVincenzo, C. Fuchs, T. Mor, E. Rains, P. Shor, J. Smolin, and W. Wootters (1999), Quantum non-locality without entanglement, *Phys. Rev. A* **59**, 1070–1091.

Brunner, N., N. Gisin, and V. Scarani (2005), Entanglement and non-locality are different resources, *New J. Phys.* **7**, 88.

Bub, J. (1977), Von Neumann's projection postulate as a probability conditionalization rule in quantum mechanics, *J. Philos. Logic* **6**, 381–390.

(2004), Why the quantum?, *Studies History Philos. Mod. Phys.* **35**, 241–266.

Clauser, J., M. Horne, A. Shimony, and R. Holt (1969), Proposed experiment to test local hidden-variable theories, *Phys. Rev. Lett.* **23**, 880–884.

Clifton, R., J. Bub, and H. Halvorson (2003), Characterizing quantum theory in terms of information-theoretic constraints, *Foundations Phys.* **33**, 1561–1591.

Cushing, J. and E. McMullin (eds.), (1989), *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem* (Notre Dame: University of Notre Dame Press).

Deutsch, D. (1997), *The Fabric of Reality* (New York: The Penguin Press).

Deutsch, D. and P. Hayden (2000), Information flow in entangled quantum systems, *Proc. Roy. Soc. London A*, **456**, 1759–1774.

Duwell, A. (2007), The many-worlds interpretation and quantum computation, *Philos. Sci.* **74**, 1007–1018.

(2008), Quantum information does exist, *Studies History Philos. Mod. Phys.* **39**, 195–216.

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

Fine, A. (2008), The Einstein–Podolsky–Rosen argument in quantum theory, in *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, E. N. Zalta (ed.), URL <http://plato.stanford.edu/archives/fall2008/entries/qt-epr/>.

Fuchs, C. (2002), Quantum mechanics as quantum information (and only a little more), arXiv:quant-ph/0205039v1.

Gisin, N. (1991), Bell's inequality holds for all non-product states, *Phys. Lett. A* **154**, 201–202.

Hardy, L. (2001), Quantum theory from five reasonable axioms, arXiv:quant-ph/0101012v4.

Howard, D. (1990), 'Nicht sein kann was nicht sein darf,' or the prehistory of EPR, 1909–1935: Einstein's early worries about the quantum mechanics of composite systems, in *Sixty-Two Years of Uncertainty: Historical, Philosophical, and Physical Inquiries into the Foundations of Quantum Mechanics*, A. Miller (ed.) (New York: Plenum), pp. 61–111.

Horodecki, M., P. Horodecki, and R. Horodecki (1998), Mixed-state entanglement and distillation: is there a 'bound' entanglement in nature? *Phys. Rev. Lett.* **80**, 5239–5242.

Jaeger, G. (2007), *Quantum Information: An Overview* (New York: Springer).

(2009), *Entanglement, Information, and the Interpretation of Quantum Mechanics* (New York: Springer).

Jarrett, J. (1984), On the physical significance of the locality conditions in the Bell arguments, *Noûs* **18**, 569–589.

Maudlin, T. (1994), *Quantum Nonlocality and Relativity* (Oxford: Blackwell).

Mayers, D. (1997), Unconditionally secure quantum bit commitment is impossible, *Phys. Rev. Lett.* **78**, 3414–3417.

Niset, J. and N. Cerf (2006), Multipartite non-locality without entanglement in many dimensions, *Phys. Rev. A* **74**, 052103.

Penrose, R. (1998), Quantum computation, entanglement and state reduction. *Philos. Trans. Roy. Soc. London A* **356**, 1927–1939.

Popescu, S. (1995), Bell's inequalities and density matrices: revealing 'hidden' non-locality, *Phys. Rev. Lett.* **74**, 2619–2622.
Popescu, S. and D. Rohrlich (1994), Nonlocality as an axiom, *Foundations Phys.* **24**, 379–385.
Schrödinger, E. (1935a), Discussion of probability relations between separated systems, *Proc. Cambridge Philos. Soc.*, **31**, 555–562.
Schrödinger, E. (1935b), Die gegenwärtige Situation in der Quantenmechanik, *Naturwissenschaften*, **23**, 807–812, 823–828, 844–849; English translation by J. D. Trimmer (1980), The present situation in quantum mechanics: a translation of Schrödinger's 'cat paradox' paper, *Proc. Am. Philos. Soc.* **124**, 323–338.
Schumacher, B. (1995), Quantum coding, *Phys. Rev. A* **51**, 2738–2747.
Shannon, C. (1948), A mathematical theory of communication, *Bell Syst. Techn. J.* **27**, 379–423, 623–656.
Shimony, A. (1984), Controllable and uncontrollable non-locality, in *Foundations of Quantum Mechanics in Light of the New Technology*, S. Kamefuchi *et al.* (eds.) (Tokyo: Physical Society of Japan), pp. 225–230. (Reprinted in A. Shimony (1993), *Search for Naturalistic World View*, Vol. 2 (Cambridge: Cambridge University Press), pp. 130–139.)
(1995), Degree of entanglement, *Annals New York Acad. Sci.* **755**, 675–679.
(2008), Bell's theorem, in *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, E. N. Zalta (ed.), URL <http://plato.stanford.edu/archives/fall2008/entries/bell-theorem/>.
Shor, P. (1994), Algorithms for quantum computation: discrete log and factoring, in *Proceedings of the 35th Annual Symposium on the Foundations of Computer Science*, S. Goldwasser (ed.) (New York: IEEE Computer Society Press), pp. 124–134.
Timpson, C. (2006), The grammar of teleportation, *Brit. J. Philos. Sci.* **57**, 587–621.
Werner, R. (1989), Quantum states with Einstein–Podolsky–Rosen correlations admitting a hidden-variable model, *Phys. Rev. A* **40**, 4277–4281.
Werner, R. and M. Wolf (2001), Bell inequalities and entanglement, arXiv:quant-ph/0107093v2.
Wheeler, J. (1990), Information, physics, quantum: the search for links, in *Complexity, Entropy and the Physics of Information*, W. Zurek (ed.) (Redwood City, CA: Addison-Wesley), pp. 3–28.
Zeilinger, A. (1999), A foundational principle for quantum mechanics, *Foundations Phys.* **29**, 631–643.

Part I

Quantum entanglement and non-locality

1

Non-locality beyond quantum mechanics

Sandu Popescu

For Abner Shimony. Your influence on me goes well beyond physics. Knowing you and being close to you is one of the greatest privileges and pleasures in my life.

1.1 Introduction

Quantum mechanics is, without any doubt, a tremendously successful theory: it started by explaining black-body radiation and the photoelectric effect, it explained the spectra of atoms, and then went on to explain chemical bonds, the structure of atoms and of the atomic nucleus, the properties of crystals and the elementary particles, and a myriad of other phenomena. Yet it is safe to say that we still lack a deep understanding of quantum mechanics – surprising and even puzzling new effects continue to be discovered with regularity. That we are surprised and puzzled is the best sign that we still don't understand; however, the veil over the mysteries of quantum mechanics is starting to lift a little.

One of the strangest things microscopic particles do is to follow non-local dynamics and to yield non-local correlations. That particles follow non-local equations of motion was discovered by Aharonov and Bohm [1], while non-local correlations – which are the subject of this chapter – were discovered by John Bell [2] and first cast in a form that has physical meaning, i.e., that can be experimentally tested, by Clauser, Horne, Shimony, and Holt [3]. When they were discovered, both phenomena seemed to be quite exotic and at the fringe of quantum mechanics. By now we understand that they are some of the most important aspects of quantum-mechanical behavior.

Consider two experimentalists, Alice and Bob, situated on different planets, far from each other. They perform experiments on particles that come from a common source and are prepared in a so-called entangled state. The experiments are "space-like separated." They take a short time compared with the time required for light, and, according to Einstein, for any other signal, to propagate from Alice to Bob.

Philosophy of Quantum Information and Entanglement, ed. A. Bokulich and G. Jaeger. Published by Cambridge University Press.

Furthermore, by pre-arrangement, the experiments are timed in such a way that Alice's experiment finishes before she could receive any signal from Bob about the experiment he performed (what experiment he did and what the result was) and similarly all information from Alice about the experiment she performed can reach Bob only after he has finished his experiment. Nevertheless, their results turn out to be correlated (although this can be found out only later, when Alice and Bob are able to compare their results). The fact that the results are correlated is not a great surprise – after all, the particles came from a common source. What is astonishing, however, is that they are correlated in such a way that, if we want to establish such correlations with any classical devices, they have to communicate with each other. Owing to the timing of the experiments, this communication has to be superluminal. We refer to such correlations as non-local.

That non-local correlations can exist at all, and not lead immediately to conflict with Einstein's relativity, is possible only because the outcomes of the measurements are probabilistic. It is the fact that quantum mechanics is fundamentally indeterministic that opens an umbrella under which non-locality can "peacefully coexist," as Abner Shimony said, with relativity [4]. This is true not only for non-local correlations but also for the non-local equations of motion discussed by Aharonov and Bohm.

While non-locality requires indeterminacy, one can easily imagine indeterministic theories that do not present non-locality. This led Aharonov [5] and Shimony [6] independently to suggest that non-locality is a deeper aspect than indeterminism, and may be the reason why quantum mechanics is what it is. In effect, they suggested that relativity and the existence of non-locality could be the axioms that determine quantum mechanics.

Following these suggestions, Daniel Rohrlich and I asked whether quantum mechanics is the only possible theory that allows the coexistence of non-locality and relativity [7]. As a first step we asked whether there could be non-local correlations that do not violate relativity (are "non-signaling") but cannot be obtained from quantum mechanics. To our surprise we found that such correlations are theoretically possible. But are there such correlations in nature? If yes, where? And if not, why not?

While, after asking the question, finding out about the theoretical possibility of non-local correlations stronger than those arising from quantum mechanics proved to be relatively easy, understanding the significance of this discovery is far more difficult. Here I will describe some steps toward this goal. The story presented here is just a small part of the research in this direction. After its discovery, the idea of non-locality beyond quantum mechanics lay dormant for more than a decade (despite a breakthrough by van Dam [8], which was not published and hence went largely unnoticed). The subject was revived by Barrett *et al.* [9], who established a

framework for describing these correlations. At present the study of these correlations is a very active research area [10].

1.2 Non-local correlations beyond quantum mechanics

What allows one to derive general statements about the nature of correlations is the fact that experiments such as those of Alice and Bob can be described in a very general, model-independent way. For our present purpose, a very convenient way to view experiments is as input/output devices. Alice has a black box that accepts as input a number x and yields as output a number a. Similarly, Bob has a black box that accepts an input y and yields the output b. One can think of these black boxes as entire automated laboratories, containing particles, measuring devices, computers, etc. The laboratories are pre-arranged, ready to perform a number of different experiments. The inputs x and y simply indicate which experiment is to be performed, while the outputs a and b are the results of the experiments.

We consider that, as a matter of principle, Alice and Bob cannot look inside the labs and see exactly how the outputs are obtained. Furthermore, all Alice and Bob can do is to give the inputs – they do not control the outputs. In particular, when for a given input different outputs are possible, Alice and Bob cannot force a particular output. This rule mimics the quantum-mechanical behavior – when an experiment can yield different outcomes, the experimentalist cannot control which particular outcome will be obtained. We say that such boxes exhibit "fundamental indeterminacy."

Given the above setting, the entire physics is encapsulated in $P(a, b|x, y; t_x, t_y)$, the conditional joint probabilities of obtaining the outcomes a and b when the inputs are x, given at time t_x, and y, given at time t_y.

Throughout this text we are interested only in boxes that are consistent with relativity. In our context this implies two things. First, that as long as the inputs are outside the light-cone of each other, the probabilities are independent of the exact timing, hence they reduce to $P(a, b|x, y)$. Second, that they obey non-signaling constraints, namely that the probability that Alice obtains a particular outcome a is independent from Bob's input and vice versa:

$$\begin{aligned} \sum_b P(a, b|x, y) &= P(a|x), \\ \sum_a P(a, b|x, y) &= P(b|y). \end{aligned} \tag{1.1}$$

The situation is illustrated in Figure 1.1.

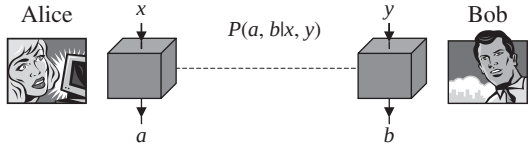


Fig. 1.1 An experimental setup viewed in terms of black boxes.

The case famously considered by Clauser, Horne, Shimony, and Holt is the simplest non-trivial case, in which all the inputs and outputs are binary. For our purposes we find it convenient to associate with x, y, a, and b the values of 0 and 1. Suppose Alice and Bob input at random, with equal probability, the values 0 and 1. It is easy to see that, in this language, what CHSH did was to analyze the probability with which the boxes succeed in yielding outputs such that

$$a \oplus b = xy, \tag{1.2}$$

where $\oplus$ denotes addition modulo 2. In other words, when x and y are both equal to 1 we succeed if the outputs are different, while in all other cases success is defined by the outputs being the same. The CHSH inequality tells us that, if the boxes work according to classical physics, the probability of success is bounded by

$$P_{\text{success}}^{\text{classical}}(a \oplus b = xy) \leq \frac{3}{4}. \tag{1.3}$$

On the other hand, if the boxes work according to quantum mechanics they can yield a larger success probability,

$$P_{\text{success}}^{\text{quantum}}(a \oplus b = xy) \leq \frac{2+\sqrt{2}}{4}. \tag{1.4}$$

(All one needs to do to find the above expressions is to convert the expectation values in the standard form of the CHSH inequality into probabilities and to note that each pair of inputs occurs with probability 1/4.)

The question Daniel Rohrlich and I asked was whether an even larger success probability could be obtained, that would be consistent with non-signaling. In the above language the answer is trivial, non-signaling doesn't constrain the maximal value of the success probability at all:

$$P_{\text{success}}^{\text{super-quantum}}(a \oplus b = xy) \leq 1. \tag{1.5}$$

Indeed, note that to each value of the product xy there correspond two different solutions for a and b: for $xy = 0$, the two solutions are $a = 0$, $b = 0$ and $a = 1$, $b = 1$, while for $xy = 1$ we have $a = 0$, $b = 1$ and $a = 1$, $b = 0$. As long as the devices yield each of the two solutions with equal probabilities, the local

probabilities for all outcomes are $1/2$ regardless of the inputs, so the correlations are non-signaling.

1.3 Communication complexity

Abner Shimony pointed out to me, and I fully agreed, that the correlations Daniel Rohrlich and I discovered are just a stand-alone example, and in order for their meaning to be evaluated they need to be integrated into a full theory that will explain all the known phenomena, along with instances in which stronger-than quantum correlations would appear. While this is certainly true, little did we know just how much one can milk this extremely simple example. The shock came for me when Wim van Dam showed me his results on communication complexity.

Consider again Alice and Bob. One day, overwhelmed by their "passion at a distance," they decide to meet for the first time. But they are very busy, so finding a day when they are both free is a difficult task. To make things more fun, they decide that, instead of trying to find a good day directly, they should first find out whether the total number of convenient days when they are both free this year is even or odd. Suppose furthermore that it is only Bob who will send information to Alice, and it is Alice who has the task of finding the result.

Of course, the task can be accomplished if Bob sends Alice his entire schedule. But this is very redundant: all Alice wants is one bit of information, i.e., "even" or "odd," but Bob has to send N bits of information, a "free" or "busy" for each of the N days of the year. Of course, they could try some other communication strategy, for example Bob could first tell Alice whether the total number of his free days is even or odd, then some other information, etc., etc. Unfortunately, it is quite easy to see that there is no method that requires less communication than sending his entire schedule.

In mathematical terms, any communication problem in which the answer is a single bit (i.e., a variable that can take only the value 0 or 1) can be formulated as follows. Alice and Bob each have N bits; Alice has $x_1, \ldots, x_N$ and Bob has $y_1, \ldots, y_N$; Alice doesn't know Bob's bits and Bob doesn't know Alice's. Let $f(\mathbf{x}, \mathbf{y})$ be a function of $\mathbf{x}$ and $\mathbf{y}$, where f can take only the value 0 or 1 and where by $\mathbf{x}$ and $\mathbf{y}$ we denote the sets of N bits $\mathbf{x} = \{x_1, \ldots, x_N\}$ and $\mathbf{y} = \{y_1, \ldots, y_N\}$. Alice and Bob know in advance the function f, and the task is for them to collaborate such that in the end Alice will find out the value of $f(\mathbf{x}, \mathbf{y})$.

To solve the problem, in general, Alice and Bob will have to communicate. We assume that they agree in advance on a specific communication protocol.

There are different problems we may consider – we are interested here in a one-way communication problem, in which it is only Bob who sends information to Alice.

Obviously, any such task can be accomplished if Bob tells Alice all his N bits. But, depending on the specific form of the function f, he might not need to send so much information. For example, if f is independent of $\mathbf{y}$ then Bob doesn't need to tell Alice anything; if f depends only on p, the parity of Bob's bits ($p = y_1 \oplus y_2 \oplus \ldots \oplus y_N$, where by $\oplus$ we denote addition modulo 2), then Bob need only send p, a single bit of information. The basic question is to find the most efficient protocol, i.e., the protocol in which Bob needs to communicate the minimum number of bits. This minimum number represents the *complexity of the communication*. Incidentally, the absolute minimum is 1 bit of communication for all cases except when the function is independent of $\mathbf{y}$. Indeed, since Alice doesn't have all the information to start with, she needs to learn at least 1 bit. Furthermore, since Alice is interested only in finding out a single bit – the value of f – any communication of more than 1 bit is redundant.[1] Finding the best protocol for an arbitrary function is, in general, a very difficult task, and it is a problem considered in computer science.

In the particular dating problem with which we started our discussion Alice and Bob need to evaluate the so-called *inner product* of $\mathbf{x}$ and $\mathbf{y}$, that is

$$f(\mathbf{x}, \mathbf{y}) = \mathbf{xy} = x_1 y_1 \oplus x_2 y_2 \oplus \ldots \oplus x_N y_N, \tag{1.6}$$

where x_i describes whether Alice's day i is busy ($x_i = 0$) or free ($x_i = 1$) and similarly y_i describes Bob's days. Day i is convenient for both Alice and Bob only if the product $x_i y_i = 1$.

As we mentioned before, evaluating the inner product is a very demanding task – Bob needs to send all his bits. The proof is extremely simple. Among all the possible patterns of free and busy days, a particular case is when Alice is free only on day 1, i.e., $x_1 = 1$, $x_2 = x_3 = \ldots = x_N = 0$. In that case $\mathbf{xy} = y_1$. So if Alice is to know this value she must know y_1. However, Alice might be free only on day 2 ($x_2 = 1$, $x_1 = x_3 = \ldots = x_N = 0$). In that case $\mathbf{xy} = y_2$. So if Alice is to know this value she must know y_2 and so on. But since Alice is not allowed to tell Bob anything, Bob doesn't know whether Alice needs to know y_1 or $y_2 \ldots$ or y_N, so he needs to send them all.

But what if Alice and Bob also share pairs of entangled particles? After all, entangled particles generate correlations that cannot be generated by any classical means, and the proof considered only classical manipulations. At first sight the idea that entanglement might help seems hopeless. Indeed, the first thing one learns about the non-local correlations generated by entangled particles is that they cannot be used for sending information. The reason for this is that these correlations

[1] Here, for simplicity, by bit we mean a binary digit – a "0" or a "1" – not the concept of a bit as a quantity of information, which takes into account the probabilities for the digit to be 0 or 1.

are established instantaneously, immediately when Alice and Bob perform measurements on their particles, and if the correlations could send information this would imply superluminal signaling. However, perhaps they might be useful *in conjunction with classical communication*, that is, in protocols that involve both non-local correlations and classical communication. Waiting for the classical information to arrive makes the whole protocol work at speeds slower than light. In fact Cleve and Buhrman [11] found that quantum entanglement can indeed help in some communication problems. However, it was also shown by Cleve *et al.* [12] that quantum-mechanical non-local correlations cannot help Alice and Bob in their dating task.

That quantum mechanics cannot help is unfortunate indeed, because Alice and Bob's dating problem (the "inner-product" problem) is not just a silly game. In fact every communication problem in which Alice and Bob want to learn just one single bit of information can be reduced to this particular problem. Thus, if we could succeed in reducing the redundancy in communication for the inner-product problem, we would reduce it in all communication problems.

What van Dam has shown is that, if Alice and Bob were to have access to "maximally" super-quantum non-local correlations, i.e., the correlations that reach the upper bound in (1.5), $P_{\text{success}}^{\text{super-quantum}}(a \oplus b = xy) = 1$ (so-called PR correlations or PR boxes), then they could solve their dating problem by using a single bit of communication, hence completely eliminating the redundancy in communication. Furthermore, since, as we mentioned before, any other communication problem in which Alice is interested only in learning 1 bit can be mapped into this dating problem, it implies that the existence of PR correlations would result in completely eliminating the redundancy of all communication.

Van Dam's solution is extremely simple. Recall that a pair of PR boxes are non-signaling input–output devices that obey the rule $xy = a \oplus b$, (1.5). To solve the inner-product problem, Alice and Bob use N PR-box pairs. They input x_1 and y_1 in the first pair, x_2 and y_2 in the second, and so on. Then the inner product is related to the outcomes of their boxes by

$$\mathbf{xy} = x_1 y_1 \oplus \ldots \oplus x_N y_N = a_1 \oplus b_1 \oplus \ldots \oplus a_N \oplus b_N. \tag{1.7}$$

By regrouping the terms in the last equality we obtain

$$\mathbf{xy} = a \oplus b, \tag{1.8}$$

where

$$\begin{aligned} a &= a_1 \oplus \ldots \oplus a_N, \\ b &= b_1 \oplus \ldots \oplus b_N. \end{aligned} \tag{1.9}$$

Hence all Alice needs from Bob is a single bit, the value of b.

The maximal non-local correlation used by van Dam is a very particular correlation and a very extreme case. Brassard *et al.* [13] made the next breakthrough: they suggested that perhaps every PR box pair with a probability of success above the quantum limit of $(2+\sqrt{2})/4 \approx 0.85$ leads to eliminating all the redundancy in communication. Using error-corrections methods they succeeded in showing that all PR boxes with success probability above approximatively 0.91 lead to eliminating redundancy in communication. Therefore, at present there is still a gap, from 0.85 to 0.91, about which we don't know anything.

1.4 Non-local computation

If the status of super-quantum correlations with respect to communication complexity is still unknown, there is another problem, namely non-local computation [14], where the boundary between quantum and super-quantum correlations is sharp.

Consider an ordinary computation problem with N input bits, $z_1, \ldots, z_N$, and a one-bit output, $c = f(z_1, \ldots, z_N)$. To this problem there corresponds a non-local version: the computation is carried out by two devices, one at Alice and one at Bob. Each device has N bits of input and an output of one bit, $x_1, \ldots, x_N$ and a, respectively, for Alice, and $y_1, \ldots, y_N$ and b for Bob. Alice and Bob are given the input bits by some external agents, such that the parity of their inputs equals the original input

$$x_i \oplus y_i = z_i. \tag{1.10}$$

For every possible value of the original input z_i there are two possible combinations of x_i and y_i that obey (1.10): when $z_i = 0$ the two combinations are $x_i = 0$, $y_i = 0$ and $x_i = 1$, $y_i = 1$, while for $z_i = 1$ we can have $x_i = 0$, $y_i = 1$ or $x_i = 1$, $y_i = 0$. The rule of the game is that for each input bit of the original problem, z_i, we give Alice and Bob one of the two corresponding sets of inputs x_i and y_i at random, with equal probability. Consequently, seeing only their own inputs, x_i and y_i, respectively, Alice and Bob have no knowledge about the original input z_i.

The task of Alice and Bob is to output one bit each, a and b, respectively, such that their parity equals the result of the original computation

$$a \oplus b = c = f(z_1, \ldots, z_N). \tag{1.11}$$

Again, for each value of c there are two possible combinations of a and b; we do *not* impose any restriction on which combination should occur. The setup for non-local computation is illustrated in Figure 1.2.

Alice and Bob know in advance the function they have to compute and they are allowed to communicate in advance, set a common strategy, and prepare their

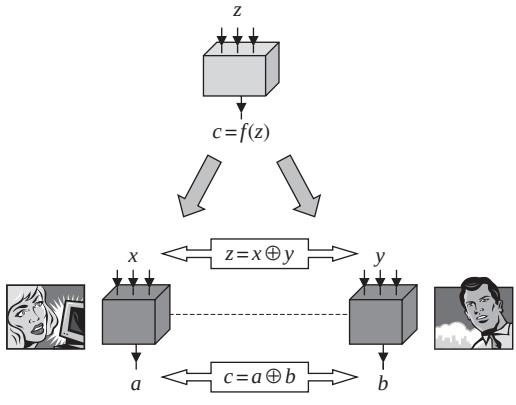


Fig. 1.2 Non-local computation.

devices in whichever way they want. However, they are no longer allowed to communicate once they have been given their inputs. To ensure this, we arrange that the whole procedure performed by Alice, from the moment when she receives her inputs until she delivers her output, is space-like separated from Bob's procedure.

There are three main cases of interest: when Alice and Bob have only classical devices, when they also use entangled quantum particles, and when they have access to super-quantum non-local correlations.

In general Alice and Bob cannot always succeed at outputting the correct answer; their task is to try to do as well as possible. There are various measures of success. A simple scenario is when Alice and Bob are given inputs at random, with equal probability, and they try to obtain the best average success probability.

In effect the whole setting we described above is a Bell-inequality experiment in which Alice can perform one out of 2^N experiments, each with a binary output, and similarly for Bob. Indeed, each of the 2^N possible combinations of, say, Alice's input bits $x_1, \ldots, x_N$ denotes an experiment she performs on her device. The 2^N experiments may all be different from each other, or some of them may be the same. The only difference from a usual Bell-inequality experiment is that Alice and Bob are given their settings by some external party, to avoid Alice and Bob cheating. The upper bound on the average probability of success when the devices are classical, $P_{\mathrm{C}}^{\max}$, is the generalized Bell inequality, whereas the upper bound on the average probability of success when the devices are quantum, $P_{\mathrm{Q}}^{\max}$, is the generalized Cirel'son inequality [15].

The astonishing result is that neither classical physics nor quantum mechanics allows non-local computation, but the very moment we go beyond the quantum-mechanical limit non-local computation becomes possible. An example suffices.

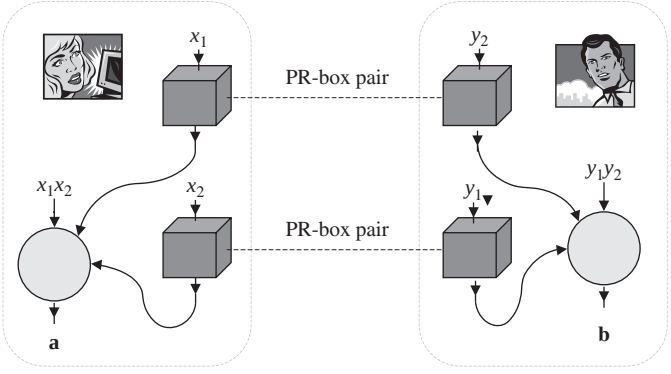


Fig. 1.3 A strategy for computing non-local-AND with two PR-box pairs.

Consider the circuit illustrated in Figure 1.3. When the two PR-box pairs are perfect this circuit succeeds, with probability 1, at computing the non-local version of the AND function of two bits. The AND function is $f(z_1, z_2) = z_1 z_2$ and its non-local version is

$$f(x_1\oplus, y_1, x_2 \oplus y_2) = (x_1 \oplus y_1)(x_2 \oplus y_2). \tag{1.12}$$

Indeed, (1.12) can be written as

$$f(x_1\oplus, y_1, x_2 \oplus y_2) = x_1x_2 \oplus y_1y_2 \oplus x_1y_2 \oplus x_2y_1, \tag{1.13}$$

where the first two terms on the right-hand side are local and the last two terms can be evaluated by PR boxes. On the other hand, it is very easy to see that the maximal probability of success obtainable by classical devices is 3/4. Indeed, the best classical strategy is for Alice's device to contain a list of pre-prepared outputs for each of the four input possibilities ($\{x_1 = 0, x_2 = 0\}$, $\{x_1 = 0, x_2 = 1\}$, $\{x_1 = 1, x_2 = 0\}$, and $\{x_1 = 1, x_2 = 1\}$) and similarly for Bob; one can immediately verify that no assignment can succeed in more than three out of four cases. Now let the PR boxes be noisy. Whenever they are still above the quantum limit (i.e., probability of success larger than $(2 + \sqrt{2})/4$) the probability of success for the complete circuit is larger than 3/4. The very moment that the PR boxes reach the quantum limit, the circuit reaches the classical limit of 3/4.

Furthermore, it is important to note that the limit of 3/4 effectively means that non-linear computation is impossible. Indeed, 3/4 is the "best-linear-approximation" limit, defined as follows. For every function $f(z_1, \ldots, z_N)$ that takes only the values 0 and 1, we can associate a "best linear approximation," that is, a function

$$g(z_1, \ldots, z_N) = \alpha_1 z_1 \oplus \alpha_2 z_2 \oplus \ldots \oplus \alpha_N z_N \oplus \beta, \tag{1.14}$$

where $\alpha_1, \ldots, \alpha_N$ and β are constants equal to 0 or 1 chosen such that g coincides with f for the maximal possible number of input values. For example $g(z_1, z_2) = 0$ is a best linear approximation for the non-linear function AND, i.e., for $f(z_1, z_2) = z_1 z_2$; in this case g coincides with f for three out of four inputs (all inputs except $z_1 = z_2 = 1$). The non-local version of g has the form

$$g(x_1, y_1, \ldots, x_N, y_N) = \alpha_1(x_1 \oplus y_1) \oplus \ldots \oplus \alpha_N(x_N \oplus y_N) \oplus \beta. \qquad (1.15)$$

Such a linear non-local function can trivially be computed perfectly even with classical devices. What was shown in [14] is that neither classical nor quantum devices can do better than this, i.e., an optimal non-local computation protocol with classical or quantum devices is simply to compute the best linear approximation. In the particular case of the AND function discussed above, the best strategy is for Alice and Bob's devices to always output 0; in this case the success probability equals 3/4. Since non-linearity is considered to be the core of computation, we can say that neither classical nor quantum devices can perform non-local computation. On the other hand, as shown in the example above, the moment we have access to some appropriate super-quantum resources (even infinitesimally stronger than the quantum ones) non-local computation becomes possible.

1.5 Conclusions

When we discovered that super-quantum non-local correlations consistent with relativity could theoretically exist, the first thought was of finding reasons why they should not exist in nature. There is no evidence that such correlations exist and, furthermore, quantum mechanics, which forbids them, is a theory that has succeeded in explaining a tremendously large range of phenomena. It is hard to imagine that one can find an alternative theory whose predictions would coincide with those of quantum mechanics in all the places where quantum mechanics has experimentally been verified, yet allow for such a radical departure as super-quantum correlations. In other words, the quest was for finding a new and natural axiom that would forbid such correlations. After years of making no progress in this quest, the attitude advocated in [8] and [9], and then followed in the intensive present research, is different: rather than agonizing over finding reasons why super-quantum correlations (apparently) don't exist in nature, one should simply take these correlations seriously, and see what one could do with them if they exist. Understanding their power helps us to put into better perspective the power and limitations of the more restricted quantum-mechanical correlations, and might ultimately lead us to find the reasons why they don't exist, if they actually don't.

Coming back to the particular stories presented here, we found out that communication in both classical and quantum mechanics is redundant, while certain super-quantum correlations are strong enough to make communication non-redundant. It is even possible that all super-quantum correlations make communication non-redundant. If super-quantum correlations don't exist in nature, is this the reason why they don't? Does nature want communication to be redundant? We also saw that neither quantum nor classical mechanics allows non-local computation, while super-quantum correlation does allow it. Again, is this a good enough reason for super-quantum correlation not to exist? Or do super-quantum correlations actually exist and all these marvelous things are possible? The search is still on.

On the way many lessons about communication, computation, cryptography, etc. have already been learned.

References

[1] Y. Aharonov and D. Bohm, *Phys. Rev.* **115**, 485 (1959).
[2] J. Bell, *Physics* **1**, 195 (1964).
[3] J. F. Clauser, M. Horne, A. Shimony, and R. A. Holt, *Phys. Rev. Lett.* **23**, 880 (1969).
[4] A. Shimony, *Search for a Naturalistic World View*, Vol. II (Cambridge: Cambridge University Press, 1993), Ch. 4.
[5] Y. Aharonov, personal communication.
[6] A. Shimony, *Proceedings of the International Symposium Foundations of Quantum Mechanics* (Tokyo: Physical Society of Japan, 1983), p. 225.
[7] S. Popescu and D. Rohrlich, *Foundations Phys.*, **24**, 379 (1994).
[8] W. van Dam, PhD Thesis, University of Oxford, 1999; quant-ph 0501159.
[9] J. Barrett, N. Linden, S. Massar, S. Pirionio, S. Popescu and D. Roberts, *Phys. Rev. A* **71**, 022101 (2005), pra/71/022101.
[10] N. J. Cerf, N. Gisin, S. Massar, and S. Popescu, *Phys. Rev. Lett.* **94**, 220403 (2005); N. Brunner, N. Gisin, S. Popescu, and V. Scarani, *Phys. Rev. A* **78**, 052111 (2008); M. Forster and S. Wolf, *Ninth International Conference on QCMC*, pp. 117–120 (New York: AIP, 2009); N. S. Jones and L. Masanes, *Phys. Rev.* **72**(5), 052312 (2005); F. Dupuis, N. Gisin, A. Hassidim, A. Methot, and H. Pilpel, *Math. Phys*. **48**, 082107 (2007); A. J. Short, S. Popescu, and N. Gisin, *Phys. Rev. A* 73, 012101 (2006); J. Barrett, L. Hardy, and A. Kent, *Phys. Rev. Lett.* **95**, 010503 (2005); M. Barnett, F. Dowker, and D. Rideout, *J. Phys. A* **40**, 7255 (2007); J. Barrett and S. Pironio, *Phys. Rev. Lett.* **95**, 140401 (2005); S. Marcovitch, B. Reznik, and L. Vaidman, *Phys. Rev. A* **75**, 022102 (2007); V. Scarani, *Proceedings of the Conference "On the Present Status of Quantum Mechanics,"* Losinj, September 2005; and *Festschrift for the 70th Birthday of GianCarlo Ghirardi, AIP Conference Proceedings*, Vol. 844, pp. 309–320 (New York: Melville, 2006); A. Broadbent and A. A. Methot, *Theor. Computer Sci.* C **358**, 3–14 (2006); S. Wolf and J. Wullschleger, *Proceedings of the International Symposium on Infinity Theory (ISIT)* (2005); J. Barrett, *Phys. Rev. A* **75**, 032304 (2007); I. Pitowsky, *Phys. Rev. A* **77**, 062109 (2008).
[11] R. Cleve and R. Buhrman, *Phys. Rev. A*, **56**, 1201 (1997).

[12] R. Cleve, W. van Dam, M. Nielsen, and A. Tapp, *Lect. Notes Computer Sci.* **1509**, 61 (1999).
[13] G. Brassard, H. Buhrman, N. Linden *et al.*, *Phys. Rev. Lett.* **96**, 250401 (2006).
[14] N. Linden, S. Popescu, A. Short, and A. Winter, *Phys. Rev. Lett.* **99**, 180502 (2007).
[15] B. Cirel'son (Tsirelson), *Lett. Math. Phys.* **4**, 93 (1980).

2

Entanglement and subsystems, entanglement beyond subsystems, and all that

Lorenza Viola and Howard Barnum

2.1 Introduction

The first realization that the validity of the quantum superposition principle in the Hilbert space describing a composite quantum system may give rise to fundamentally new *correlations* between the constituent subsystems came in the landmark 1935 paper by Einstein, Podolsky, and Rosen (EPR) [1], where it was shown how the measurement statistics of observables in certain quantum states could not be reproduced by assigning definite wavefunctions to individual subsystems. It was in response to the EPR paper that Schrödinger, in the same year, coined the term *entanglement* (*Verschränkung*) to acknowledge the failure of classical intuition in describing the relationship between the "parts" and the "whole" in the quantum world [2]:

> Whenever one has a complete expectation catalog – a maximum total knowledge – a ψ function – for two completely separated bodies, . . . then one obviously has it also for the two bodies together. But the converse is not true. The best possible knowledge of a total system does not necessarily include total knowledge of all its parts, not even when these are fully separated from each other and at the moment are not influencing each other at all.

While Bell's strengthening of the original EPR-paradox setting [3] and the subsequent experimental verification of Bell inequalities [4, 5] irreversibly changed the perception of entanglement from a property of counterintuitive "spookiness" to (beyond reasonable doubt) an experimental reality, the concept and implications of entanglement continue to be associated with a host of physical, mathematical, and philosophical challenges [6]. In particular, investigation of entanglement in both its *qualitative* and *quantitative* aspects has intensified under the impetus of quantum information science (QIS). Building on the discovery of the teleportation and dense-coding protocols [7, 8], entanglement has nowadays been identified as the defining resource for quantum communication [8], as well as an essential ingredient for understanding and unlocking the power of quantum computation [9].

Philosophy of Quantum Information and Entanglement, ed. A. Bokulich and G. Jaeger. Published by Cambridge University Press.

Furthermore, entanglement is gaining a growing status as a key bridging notion between QIS and various subfields of physics – most notably quantum foundations, quantum statistical mechanics, quantum gravity, and condensed-matter theory. In spite of continuous progress, however, the current state of entanglement theory is still marked by a number of outstanding unresolved problems, which range from the complete classification of mixed-state bipartite entanglement to entanglement in systems with continuous degrees of freedom, and the classification and quantification of multipartite entanglement for arbitrary quantum states.[1]

At an even more fundamental level, recent indications show that the very *definition* of entanglement as given thus far may be too restrictive to embrace relevant physical and information-theoretic settings in their full generality. From an operational standpoint, the distinction between entangled and unentangled states of a composite quantum system largely stems from having acknowledged a separation between "local" capabilities – thereby regarded as a "cheap" resource – as opposed to arbitrary "non-local" capabilities – which are not readily available, and hence come with a cost: were no operational restriction in place, then clearly all pure states of the system would be equivalent. In the conventional approach to entanglement, local capabilities and degrees of freedom are further (more or less explicitly) identified with spatially separated distinguishable subsystems. While such identification is both natural and adequate for the majority of QIS settings, compelling motivations for critically reconsidering the resulting subsystem-based notion of entanglement arise in situations where the identification of "local" resources may not be a-priori obvious or may conflict with additional or different restrictions. A most prominent example in this sense (and one that has received extensive attention in the recent literature, see, e.g., [10–13] for representative contributions) is offered by many-body systems consisting of indistinguishable (bosonic or fermionic) quantum particles. Whenever the spatial separation between the latter is small enough for quantum statistics to be important, admissible quantum states and observables are effectively constrained to lie in a proper (symmetric or antisymmetric) subspace of the respective tensor products of observable spaces, making the identification of "local" subsystems and operations far more delicate and, ultimately, ambiguous than in the standard case. So, in general, how can entanglement be understood in an *arbitrary* physical system, subject to *arbitrary constraints* on the possible operations we may perform for describing, manipulating, and observing its states?

Our proposed answer builds on the idea that *entanglement is an inherently relative concept*, whose essential features may be captured in general in terms

[1] See http://www.imaph.tu-bs.de/qi/problems for a current list of such problems.

of the relationships between different *observers* – as specified through expectations of quantum observables in different, physically relevant sets. In the simplest instance, distinguished observables in a preferred set determine the analog of restricted, "local" capabilities, as opposed to unrestricted, "global" capabilities embodied by the full observable space. *Generalized entanglement* (GE) of a quantum state *relative* to the distinguished set may then be defined without reference to a decomposition of the overall system into subsystems [14, 15]. That the role of the observer must be properly acknowledged in determining the distinction between entangled and unentangled states has been stressed by various authors in various contexts: in particular, the emergence of distinguished subsystems and of a preferred tensor-product structure has been related to the set of operationally available interactions and measurements in [16, 17], whereas the presence of maximal entanglement in a state has been directly defined in terms of maximal fluctuations of fundamental observables in [18, 19]. In spite of suggestive points of contact, our approach differs from the above in (at least) two important ways: physically, the need for a decomposition into distinguishable subsystems is bypassed altogether; mathematically, the GE notion rests directly (and solely) on *extremality properties of quantum states in convex sets* that are associated with different observers. Therefore, GE is both directly applicable to arbitrary operator subspaces and algebraic languages that may be used to specify the relevant quantum system and suitable for investigations of general operational theories, where convexity plays a key role [20, 21].

A more rigorous and thorough development of the GE framework is available in [14, 20, 22]; our main goal here is to informally revisit the key steps and further illustrate them through examples that may be especially useful at highlighting conceptual departures from the standard view.

2.2 Entanglement and subsystems: the standard view

In order to motivate and introduce the concept of GE, we begin by briefly revisiting the standard setup of entanglement theory. Throughout this paper, our formal treatment of GE properties of quantum systems will be confined to *finite-dimensional* quantum systems S, with associated state spaces $\mathcal{H}$, $\dim(\mathcal{H}) = d < \infty$. In line with Schrödinger's original definition, we furthermore assume that *best possible knowledge is available for* S, thus requiring S to be in a *pure* state $|\psi\rangle \in \mathcal{H}$. Let $\mathcal{B}(\mathcal{H})$ denote the space of linear operators on $\mathcal{H}$. The set of (traceless) quantum observables of S may be naturally identified with the (real) Lie algebra $\mathfrak{su}(d) \subset \mathcal{B}(\mathcal{H})$, with Lie bracket $[X, Y] = \mathrm{i}(XY - YX)$. Arbitrary states of S are represented by positive, normalized elements in $\mathcal{B}(\mathcal{H})$ (density operators), with $\rho = |\psi\rangle\langle\psi|$ for a pure state. The set Υ of density operators is a compact convex subset of $\mathcal{B}(\mathcal{H})$: thus

equivalently, ρ is pure if it is an *extreme* point of Υ, or the purity of ρ is maximal, $P(\rho) = \mathrm{Tr}(\rho^2) = 1$.

The essential intuition on which the GE notion builds may be appreciated starting from the simplest instance of a composite quantum system, namely a pair of two-dimensional subsystems (Alice and Bob henceforth), living in a four-dimensional complex space

$$\mathcal{H} \equiv \mathcal{H}_{\mathsf{AB}} = \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}, \tag{2.1}$$

where $\dim(\mathcal{H}_{\mathrm{A}}) = \dim(\mathcal{H}_{\mathrm{B}}) = 2$. Any joint pure state of Alice and Bob that is *separable*, that is, able to be expressed in the form

$$|\Psi\rangle_{\mathsf{AB}} = |\psi\rangle_{\mathsf{A}} \otimes |\phi\rangle_{\mathsf{B}} \equiv |\psi\phi\rangle_{\mathsf{AB}}, \tag{2.2}$$

for suitable subsystem states of Alice and Bob alone, is unentangled. Let $\{|0\rangle, |1\rangle\}$ denote an orthonormal basis in $\mathbb{C}^2$ (computational basis), as usual. Then the following *Bell states* are well known to be (maximally) entangled [3],

$$\begin{aligned} |\Psi^{\pm}\rangle_{\mathsf{AB}} &= \frac{|00\rangle_{\mathsf{AB}} \pm |11\rangle_{\mathsf{AB}}}{\sqrt{2}}, \\ |\Phi^{\pm}\rangle_{\mathsf{AB}} &= \frac{|01\rangle_{\mathsf{AB}} \pm |10\rangle_{\mathsf{AB}}}{\sqrt{2}}. \end{aligned} \tag{2.3}$$

What distinguishes, at an operational level, states of the form (2.2) from states of the form (2.3)? While the answer to this question may be phrased in different ways in principle, it ultimately rests on the distinction between what Alice and Bob may accomplish in terms of purely *local* resources as opposed to arbitrary non-local ones. Let, in particular,

$$\Omega_{\mathrm{loc}} = \mathrm{span}_{\mathbb{R}}\{A \otimes \mathbb{1}, \mathbb{1} \otimes B \mid A = A^{\dagger}, B = B^{\dagger}\} \tag{2.4}$$

denote the set of local traceless observables on $\mathcal{H}_{\mathsf{AB}}$, so that a generic unitary transformation generated by observables in Ω_{loc} is of the form $U_{\mathsf{AB}} = U_{\mathsf{A}} \otimes U_{\mathsf{B}}$. (The traceless condition excludes the identity operator from the distinguished subspace of observables; it is a somewhat arbitrary decision whether or not to do this, for it has no effect on the convex structure of the set of reduced states induced by normalized states, since all states take the same value on it. Considering the reduced states as a base for a cone, however, is equivalent to reintroducing the identity operator, and considering the reduced states induced by the states in the cone of *unnormalized* states.) Note that Ω_{loc} may be identified with the Lie algebra

$$\Omega_{\mathrm{loc}} \simeq \mathfrak{su}(2)_{\mathsf{A}} \oplus \mathfrak{su}(2)_{\mathsf{B}} = \mathrm{span}_{\mathbb{R}}\{\sigma_a \otimes \mathbb{1}, \mathbb{1} \otimes \sigma_b\}, \tag{2.5}$$

where $\sigma_a, a \in \{x, y, z\}$, denote spin Pauli operators on $\mathbb{C}^2$. Then no Bell state may be reached starting from a product state as in (2.2) solely by application of

operators generated by local observables. Alternatively, imagine that the state of Alice and Bob is to be determined solely on the basis of access to expectation values of observables.[2] Then any pure product state is completely specified by knowledge of the expectation values $\langle\sigma_a\rangle_{\mathsf{A,B}}$ on each subsystem, whereas knowledge of the same expectations *cannot* distinguish a Bell state from a *mixture* of pure states – containing no entanglement. For instance,

$$|\Phi^-\rangle_{\mathsf{AB}} \equiv_{\rm loc} \frac{|01\rangle_{\mathsf{AB}}\langle 01| + |10\rangle_{\mathsf{AB}}\langle 10|}{2}, \tag{2.6}$$

where equivalence means indistinguishability by access to expectations of restricted (local) observables. In order to distinguish, knowledge of appropriate *correlations* is required, that is expectations of non-local observables like $\sigma_x \otimes \sigma_x$ in the above example. An equivalent characterization of entanglement may be obtained in terms of purity of Alice and Bob subsystem states, as given by the corresponding reduced density operators: pure product states are precisely those states for which *each* subsystem remains pure. To state it differently, pure entangled states are those pure states whose *reduced* states – i.e., the expectation values they determine for all *one-party* observables only – are *non-extremal*, i.e., mixed.

Even in the simplest setting under consideration, it is essential to acknowledge that the characterization of a pure state in $\mathcal{H}$ as entangled or not is unambiguously defined only after a *fixed* tensor decomposition has been chosen among the distinct ones that $\mathcal{H}$ can a-priori support (as long as its dimension d is a non-prime integer). By its very nature, (standard) entanglement is *relative* to a preferred subsystem decomposition – capturing the specific way in which S is viewed as constituted of its parts. Physically, one may expect that what makes a given factorization preferred among others should be naturally linked to the set of operations which are deemed practically available for control and observation. Indeed, in the bipartite setting discussed above, the availability of arbitrary observables in $\Omega_{\rm loc}$ may be intuitively (and formally) related to the identification of local degrees of freedom associated with subsystems A and B – as described in general by appropriate mutually commuting associative algebras of operators [16, 17]. Suppose, however, that for whatever reason the rule distinguishing what is "cheap" from what is not changes – in particular, suppose that the accessible observables on $\mathcal{H} = \mathbb{C}^4$ are arbitrary linear combinations in the following set:

$$\Omega'_{\rm loc} = {\rm span}_{\mathbb{R}}\{\sigma_x \otimes \sigma_x, \sigma_z \otimes \sigma_z, \sigma_x \otimes \sigma_y, \sigma_y \otimes \sigma_z\}. \tag{2.7}$$

[2] This scenario is realistic whenever the system can be accessed only collectively in ensembles involving "unconditional" (non-selective) measurements on either a large ensemble of individual constituents or a large number of repetitions starting from identical states, see also [14].

Then expectation values of observables in $\Omega'_{\rm loc}$ are clearly sufficient to completely specify Bell states as given in (2.3), whereas pure states of the form $|\psi\phi\rangle_{\mathsf{AB}}$ may become "locally" indistinguishable from mixtures, for instance

$$\frac{|0\rangle_{\mathsf{A}} + \mathrm{i}|1\rangle_{\mathsf{A}}}{\sqrt{2}} \otimes \frac{|0\rangle_{\mathsf{B}} + \mathrm{i}|1\rangle_{\mathsf{B}}}{\sqrt{2}} \equiv_{\mathrm{loc}'} \frac{|\Psi^-\rangle_{\mathsf{AB}}\langle\Psi^-| + |\Phi^+\rangle_{\mathsf{AB}}\langle\Phi^+|}{2}, \tag{2.8}$$

where equivalence has the same meaning as before under $\Omega'_{\rm loc}$. Accordingly, Bell states should now be regarded as *un*-entangled with respect to the relevant set of local capabilities. In fact, it is possible to show in this case [16, 17] that a well-defined decomposition of $\mathcal{H}$ into subsystems still exists relative to the new local set $\Omega'_{\rm loc}$,

$$\mathcal{H} = \mathcal{H}_\chi \otimes \mathcal{H}_\lambda, \tag{2.9}$$

where χ, λ specify "virtual" subsystems corresponding to eigenvalue $\chi \in \{\Psi, \Phi\}$ and $\lambda \in \{+, -\}$, respectively – so that, for instance, $|\Psi^+\rangle \simeq |\Psi\rangle_\chi \otimes |+\rangle_\lambda$, and so on.

While the above examples nicely demonstrate how, even within the standard "subsystem-based" framework, the notion of entanglement is strongly observer-dependent, a closer scrutiny rapidly leads to the following deeper questions:

Are subsystems general and flexible enough to capture the relativity of entanglement in full? Once a preferred observer has been specified through the identification of a distinguished observable set, is it always possible to relate such an observer to the emergence of preferred subsystems? If so, is it always necessary or useful?

Several concrete examples may be adduced toward showing the inadequacy of subsystem-based entanglement in settings involving operational or fundamental constraints more general than the ones implied by the standard framework [14, 22–25]. Within the four-dimensional state space considered so far, imagine for instance that available operations are subject to a *conservation law*, say conservation of the total (pseudo)spin angular momentum along a given axis, $S_z = (s_z \otimes \mathbb{1} + \mathbb{1} \otimes s_z)/2$, with $s_a = \sigma_a/2$. Assume we describe the corresponding observer in terms of the (smallest) Lie algebra of observables commuting with S_z,

$$\Omega^z_{\rm loc} = \mathrm{span}_{\mathbb{R}}\{s_z \otimes \mathbb{1},\, \mathbb{1} \otimes s_z,\, \sqrt{2}(s_x \otimes s_x + s_y \otimes s_y),\, \sqrt{2}\,(s_x \otimes s_y - s_y \otimes s_x)\} \simeq \mathfrak{u}(2), \tag{2.10}$$

where an orthonormal Hermitian basis has been used for later reference. On the one hand, because only a subset of local observables is included, only a proper subspace of product states in $\mathcal{H}_{\mathsf{AB}}$ may expected to remain "unentangled" given such capabilities. On the other hand, although in analogy with $\Omega'_{\rm loc}$ expectations of observables in $\Omega^z_{\rm loc}$ continue to determine states like $|\Psi^\pm\rangle$, it is also clear that not all four Bell states are now on the same footing, as long as $s_z \otimes s_z$ is not included in

the distinguished set. Hence, $\Omega^z_{\rm loc}$ corresponds neither to the factorization in (2.1) nor to that in (2.9). Yet, one would still like a natural and meaningful notion of entanglement to exist on the basis of the identification of "local" resources such as the ones described by $\Omega^z_{\rm loc}$.

Aside from being desirable on fundamental grounds, compelling motivations for consistently describing a distinguished $\mathfrak{u}(2)$ observable algebra (and higher-dimensional generalizations) arise in the context of defining entanglement in systems of *indistinguishable fermions* – in which case the latter is reinterpreted in a second-quantized language and the S_z-constraint is imposed by fermion-number conservation (see Section 2.4.4). In general, quantum indistinguishability constrains the admissible fermionic states to the fully antisymmetric subspace of $\mathcal{H}$, preventing a direct identification between particles and subsystems in the standard tensor-product sense. While factorizations into distinguishable subsystems can still be defined in terms of appropriate sets of *modes*, the choice of preferred modes may be problematic in situations where different sets (e.g., spatial and momentum modes) are equally relevant to the description. Finally, the possibility that the operator *language* which describes the problem may itself be changed – for instance via isomorphic mappings between spaces of spin and of fermion operators like the Jordan–Wigner transformation [26] – further adds to the intricacy of applying the standard entanglement framework to general quantum many-body systems.

Can entanglement be *directly defined in terms of distinguished physical observables*, irrespective of and without reference to a preferred subsystem decomposition?

2.3 Entanglement beyond subsystems: the concept of generalized entanglement

The generalized entanglement setting we are seeking must satisfy two essential requirements: it must both (i) reduce to the ordinary framework in well-defined limiting situations, and (ii) identify the presence or absence of entanglement independently of the specific operator language in which the distinguished observables of S may be expressed. The key intuition is to generalize the characterization of pure-state entanglement as *relative mixedness under restricted capabilities*.

While we refer to [20] for a more detailed and mathematically more rigorous account, all relevant GE settings are subsumed as special cases of a general definition based on *distinguished cone-pairs*. Recall that a (finite-dimensional) convex cone is a proper subset of a (finite-dimensional) real vector space closed under multiplication by non-negative scalars and under addition. In our usage, the term will refer to *regular* convex cones, that is, cones that are pointed (contain no subspace other than $\{0\}$), generating (so that the linear span of the cone is the full ambient

vector space), and topologically closed. Let V and V^*, respectively, denote a real linear space and its dual, that is, the space of all linear functionals from V to $\mathbb{R}$. Given a convex cone $C \subseteq V$ and a distinguished functional $\lambda \in V^*$, we associate states with normalized elements in $x \in C$, satisfying $\lambda(x) = 1$. We require that λ separate C from $-C$, equivalently that the only element $x \in C$ for which $\lambda(x) = 0$ is $x = 0$, and we also require that $\lambda(C) \geq 0$. These conditions are imposed so that (given the regularity of the cone) the set $\hat{C}$ of normalized states in C is a *compact convex set*. Also, each element of C can be written as αx for some unique $x \in \hat{C}$, $\alpha > 0$ (in other words, $\hat{C}$ is a *base* for the cone). Thus C may be thought of as the set of *unnormalized* states: multiples of the normalized states which belong to $\hat{C}$. We will use the term *extremal state* to mean an extremal element of $\hat{C}$, in the standard sense that it cannot be written as a non-trivial convex combination of two distinct elements of $\hat{C}$. But, where it is clear that the state is in general unnormalized, we may use the term extremal state to refer to an unnormalized state belonging to an extremal ray of C, that is, a non-negative multiple of an extremal normalized state.[3]

The operational interpretation of the above construction also requires us to consider the dual cone $C^* = \{\alpha \in V^* | \alpha(x) \geq 0 \ \forall x \in C\}$, which consists of the functionals non-negative on C. C^* is interpreted as the set of possible "effects," corresponding to "unnormalized" measurement outcomes, for states in C; $\alpha(x)$, which is non-negative for all $x \in C$, is interpreted as an (unnormalized) probability for outcome α when the state is x. Observe that the distinguished functional λ (often called the "unit" or "order unit") belongs to C^*: it is the measurement outcome that has probability 1 in all states. The interval $[0, \lambda] \subset C^*$ defined by $\{\alpha \in C^* | \alpha \in C, \lambda - \alpha \in C\}$ corresponds to outcomes normalized so that they can appear in a measurement, since it is easily verified that $[0, \lambda]$ is precisely the set of functionals α such that, for all $x \in \hat{C}$, $0 \leq \alpha(x) \leq 1$, enforcing that probabilities for measurement outcomes lie in $[0, 1]$. Measurements with a finite set of outcomes correspond to finite *resolutions of the unit* into elements of $[0, \lambda]$: finite sequences α_i of elements of $[0, \lambda]$ such that $\sum_i \alpha_i = \lambda$, which, as is easily checked, enforces that the probabilities of the measurement outcomes i sum to 1 in all states, i.e., $\sum_i \alpha_i(x) = 1$. Note that for many (but not all) purposes it may be easier to ignore this additional derived structure of a dual cone and a unit interval within it, viewing V^* as simply a set of real-valued linear-in-the-states "observables" and the states in C as determining their expectation values.

Generalized entanglement may be defined once a pair of distinguished state-sets – generalizing the distinction between states as accessible to a "local" as

[3] We warn the reader that such states (0 aside) are *not* extremal in the convex-sets sense, in the convex set C.

opposed to a "global" observer – and a choice of an appropriate normalization-preserving linear map – generalizing the notion of computing the reduced density operator for bipartite systems – have been specified.

Definition 1. Let V, W be real linear spaces equipped with distinguished convex cones $C \subset V$, $D \subset W$ and positive linear functionals $\lambda \in C^*$, $\tilde{\lambda} \in D^*$ satisfying the requirements discussed above. Let $\pi\colon V \to W$ be a normalization-preserving linear map taking C onto D that is,

(i) $\pi(C) = D$, *and*
(ii) $\pi(\{x \in V | \lambda(x) = 1\}) = \{y \in W | \tilde{\lambda}(y) = 1\}$.

A pure (extremal) state $x \in C$ is *generalized unentangled relative to* D if its image $\pi(x)$ is pure (extremal) in D. A mixed (non-extremal) state in $z \in C$ is generalized unentangled relative to D if $z = \sum_a w_a x_a$, for positive numbers $w_a \geq 0$ and extreme points $x_a \in C$ whose images $\pi(x_a)$ are extreme in D.

Note that this definition applies in general to unnormalized states, but includes normalized ones. If the states are normalized, it coincides with a natural restricted definition, in which C, D are replaced by $\hat{C}$, $\hat{D}$ above, and extremality is taken in the convex-sets sense as extremality in $\hat{C}$, $\hat{D}$. Condition (ii) on the map π guarantees that this makes sense, by ensuring that $\hat{C}$ is mapped onto $\hat{D}$.

2.3.1 Generalized-entanglement settings

In physical applications, the distinguished functional λ is identified with the trace map, and the reduced state-set associated with D is obtained by selecting a preferred subspace $\Omega \subseteq \mathcal{B}(\mathcal{H})$ of "local" observables for S. In particular, the above general definition may be specialized to the following relevant entanglement settings [20].

GE1: Distinguished quantum observables setting. C is isomorphic to the convex cone Υ of quantum states on $\mathcal{H}$ (positive normalized functionals η induced by positive multiples of density operators), and V^* is viewed as the subspace of Hermitian operators in $\mathcal{B}(\mathcal{H})$, with $X[\eta] = \mathrm{Tr}(\rho_\eta X)$ giving the expectation value of observable X in state ρ_η. Elements in $V = (V^*)^*$ are arbitrary linear functionals defined so that $\xi[X] = X[\xi]$, for all $X \in V^*$, $\xi \in V$. Let the distinguished observable space be any real linear subspace $\Omega \equiv W^* \subset V^*$. Then the cone $D \subset W$ of Ω-reduced unnormalized states may be obtained as the image of C under the map $\pi\colon V \mapsto W$ which restricts elements of V to observables in the distinguished set, i.e., for $\eta \in V$, $\pi(\eta)$ is defined by the condition that, for all $X \in W^* = \Omega$, $\pi(\eta)[X] = \eta(X)$. Thus, elements in D may be seen as lists of expectation values for observables in Ω.

GE2: Hermitian-closed operator subspace setting. In this case, the distinguished observable space Ω consists of the Hermitian operators in a linear subspace $\mathcal{V} \subset \mathcal{B}(\mathcal{H})$ containing the identity and closed under Hermitian conjugation. This formulation turns out to be fully equivalent to the GE1 setting defined above.

GE3: Lie-algebraic setting. In situations where the Hilbert space $\mathcal{H}$ of S may be identified with the representation space of a (Hermitian-closed) Lie algebra $\mathfrak{h}$, it is natural to identify Ω with the corresponding cone D of reduced $\mathfrak{h}$-states consisting of linear functionals on $\mathfrak{h}$ induced from positive normalized operators in $\mathcal{B}(\mathcal{H})$ upon restriction to $W^* = \Omega$. By construction, this setting is a special case of GE1. Under the additional assumptions that $\mathfrak{h}$ is semi-simple and acts irreducibly on $\mathcal{H}$, this setting leads to an explicit characterization of the set of pure generalized unentangled states, which are identical with the set of so-called *generalized coherent states* (GCSs) of $\mathfrak{h}$ [14, 27].

GE4: Associative-algebraic setting. Here, the distinguished set Ω consists of the Hermitian elements of an associative sub-algebra $\mathcal{A}$ of $\mathcal{B}(\mathcal{H})$. This case may also be seen as a special instance of GE1. Under the additional assumption that $\mathcal{A}$ partitions into a collection of independently accessible, mutually commuting associative sub-algebras $\{\mathcal{A}_i\}$ satisfying $\otimes_i \mathcal{A}_i \simeq \mathcal{B}(\mathcal{H})$ (or generalizations where a proper "coding" subspace $\mathcal{C} \subset \mathcal{H}$ is considered), this setting is directly relevant to identifying observable-induced subsystem decompositions as in [17].

2.3.2 Generalized-entanglement measures

As with standard entanglement, no single measure can, in general, uniquely characterize the GE properties of a state. However, the observation that standard pure-state entanglement translates into mixedness (loss of purity) of the reduced subsystem states naturally suggests seeking a way to quantify the degree of purity relative to the distinguished observable set.

Let Ω be a Hermitian-closed set, and let $\{X_\alpha\}$ be an orthonormal basis of Ω, $\mathrm{Tr}(X_\alpha X_\beta) = \delta_{\alpha\beta}$. Then, for every density operator ρ on S, the projection of ρ onto Ω may be defined via

$$\mathcal{P}_\Omega(\rho) = \sum_\alpha \mathrm{Tr}(\rho X_\alpha) X_\alpha.$$

$\mathcal{P}_\Omega(\rho)$ may be regarded as the "Ω-reduced" density operator associated with ρ, although (unlike in the standard multipartite case) $\mathcal{P}_\Omega(\rho)$ is not necessarily positive semidefinite: the positivity of $\mathrm{Tr}(\mathcal{P}_\Omega(\rho)X)$ does not in general hold for all positive semidefinite observables X; although, if we include the identity in Ω, it is assured for those X belonging to Ω.

Definition 2. For any density operator $\rho \in \Upsilon$, the *purity of ρ relative to* Ω (Ω-purity) is given by the squared length of its Ω-reduced density operator, that is,

$$P_\Omega(\rho) = Tr(\mathcal{P}_\Omega(\rho)^2) = \sum_\alpha |Tr(\rho X_\alpha)|^2. \tag{2.11}$$

In particular, if $\rho = |\psi\rangle\langle\psi|$ is pure, we simply write

$$P_\Omega(|\psi\rangle) = \sum_\alpha |\langle\psi|X_\alpha|\psi\rangle|^2. \tag{2.12}$$

By construction, $0 \leq P_\Omega(\rho) \leq P(\rho) \equiv \mathrm{Tr}(\rho^2)$. Also note that $P_\Omega(|\psi\rangle) = 1/d + P_{\Omega_0}(|\psi\rangle)$, where $\Omega_0 \subseteq \Omega$ denotes the distinguished traceless sector. It is often convenient to discard the constant trace contribution, in which case the common normalization constant for the (traceless) X_α is adjusted so as to rescale the maximum of P_{Ω_0} to 1. In the Lie-algebraic setting GE3, the $\mathfrak{h}$-purity so constructed is automatically invariant under arbitrary unitary transformations in the Lie group generated by $\mathfrak{h}$, as is desirable on physical grounds.

The usefulness of the Ω-purity as a measure of *pure-state* GE in various settings is summarized in the following theorem, which is proved in [14] and [20].

Theorem. (i) In the *irreducible* Lie-algebraic and associative-algebraic settings, a pure state $|\psi\rangle \in \mathcal{H}$ is generalized unentangled relative to Ω if and only if $P_\Omega(|\psi\rangle)$ is maximal. (ii) In the Hermitian-closed operator setting (including *reducibly* represented operator algebras), states with maximal Ω-purity are generalized unentangled relative to Ω.

A complete characterization of the relationship between maximal relative purity and generalized unentanglement still remains an interesting open question in the general convex-cone framework. We refer the reader to the above-mentioned papers for an extended discussion of this issue, as well as for the construction of appropriate *mixed-state* GE measures. In what follows, we focus on developing concrete intuition about GE from several illustrative examples.

2.4 Generalized entanglement by example

2.4.1 Entanglement with respect to local observables

Assume that the system S is composed of n distinguishable subsystems, corresponding to a Hilbert-space factorization of the form

$$\mathcal{H} \simeq \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_n, \tag{2.13}$$

where for simplicity we take the factors to be isodimensional, $\dim(\mathcal{H}_\ell) = d_0$ for all ℓ, $d_0^n = d$.

2.4.1.1 The Lie-algebraic setting

Contact with the standard multipartite entanglement framework may be established by realizing that full *local accessibility* of individual subsystem states identifies the set of *arbitrary local observables* as the physically distinguished set. Let

$$\Omega_{\rm loc} \equiv \mathfrak{h}_{\rm loc} = \mathfrak{su}(d_0)_1 \oplus \ldots \oplus \mathfrak{su}(d_0)_n \tag{2.14}$$

denote the relevant (irreducible) Lie algebra of traceless local observables, generalizing the bipartite qubit case of (2.5). For each $\mathfrak{su}(d_0)_\ell$, an orthonormal (in the trace inner product) basis $\{x_\alpha\}$, $\alpha = 1, \ldots, d_0^2 - 1$, may be constructed in analogy to (normalized) Pauli operators. Thus $\mathrm{Tr}(x_\alpha x_\beta) = \delta_{\alpha,\beta}$. An overall orthonormal basis for $\mathfrak{h}_{\rm loc}$ is then obtained by extending each x_α^ℓ to act non-trivially only on subsystem ℓ, e.g.,

$$x_\alpha^{(1)} = x_\alpha \otimes \mathbb{1}^{(2)} \otimes \ldots \otimes \mathbb{1}^{(n)}, \qquad \mathbb{1}^{(\ell)} = \mathbb{1}/\sqrt{d_0}.$$

While a formal proof of the equivalence between standard entanglement and GE relative to arbitrary local observables is given in [14], the basic step follows from identifying reduced local states with lists of expectations of x_α^ℓ: because the latter completely determine reduced density operators (in the usual partial-trace sense), and a pure state $|\psi\rangle \in \mathcal{H}$ is entangled (in the standard sense) if and only if all its reduced density operators remain pure, the two notions coincide on pure states. By virtue of convexity, they can also be shown to coincide on mixed states. Consistently with this, GCSs of $\mathfrak{h}_{\rm loc}$ may be associated with orbits of a reference state like $|0 \ldots 0\rangle$ under the group of local (special) unitary transformations, $\mathrm{SU}(d_0)_1 \otimes \ldots \otimes \mathrm{SU}(d_0)_n$.

The connection with the ordinary entanglement framework may be further quantitatively appreciated by relating the local purity $P_{\mathfrak{h}} \equiv P_{\rm loc}$ to the conventional subsystem purities determined by reduced subsystem states. One finds [22]

$$P_{\rm loc}(|\psi\rangle) = \frac{d_0}{d_0 - 1}\left(\frac{1}{n}\sum_{\ell=1}^{n}\mathrm{Tr}(\rho_\ell^2) - \frac{1}{d_0}\right), \tag{2.15}$$

where ρ_l denotes as usual the reduced density operator of subsystem ℓ. Thus, $P_{\rm loc}$ is proportional to the *average* subsystem purity (the first term in the parentheses in (2.15)). For the special case of qubits ($d_0 = 2$), the local GE as quantified by $P_{\rm loc}$ is additionally simply related [23, 28] to a measure of *global multipartite entanglement* Q originally proposed in [29], $P_{\rm loc}(|\psi\rangle) = 1 - Q(|\psi\rangle)$.

2.4.1.2 A three-qubit case study

It is essential to realize that, in order for the equivalence between multipartite subsystem entanglement and GE to hold, *all* and *only* local observables must be

distinguished. A concrete example may serve to further illustrate these points. Let S consist of three qubits. The local algebra is given by

$$\Omega_1 = \mathfrak{su}(2)_1 \oplus \mathfrak{su}(2)_2 \oplus \mathfrak{su}(2)_3, \tag{2.16}$$

with corresponding local purity given by

$$P_1(|\psi\rangle) = \frac{1}{3}\sum_{\ell=1}^{3}\sum_{\alpha=x,y,z}\langle\sigma_\alpha^{(\ell)}\rangle^2.$$

This distinguishes four classes of states with different multipartite (and GE) properties: $P_1(|\psi\rangle) = 1$ on arbitrary product states, which are thus unentangled; $P_1(|\psi\rangle) = 1/3$ on so-called bi-separable states – joint states, e.g., of the form

$$|\mathsf{B}_{12}\rangle = |\mathtt{Bell}\rangle_{12} \otimes |\phi\rangle_3,$$

and similarly for states $|\mathsf{B}_{13}\rangle$ and $|\mathsf{B}_{23}\rangle$, where qubits 2 and 3 are factored out, respectively; $P_1(|\psi\rangle) = 1/9$ for $|\psi\rangle$ belonging to the so-called $\mathtt{W}$-class, e.g.,

$$|\mathtt{W}\rangle = \frac{1}{\sqrt{3}}\left(|001\rangle + |010\rangle + |100\rangle\right);$$

and $P_1(|\psi\rangle) = 0$ on the Greenberger–Horne–Zeilinger ($\mathtt{GHZ}$) class,

$$|\mathtt{GHZ}\rangle = \frac{1}{\sqrt{2}}\left(|000\rangle + |111\rangle\right),$$

identifying the latter as maximally entangled with respect to local observables.

Suppose, however, that "local" resources are redefined, so that arbitrary unitary operations on the first pair are distinguished – that is, we effectively replace Ω_1 with

$$\Omega_2 = \mathfrak{su}(4)_{12} \oplus \mathfrak{su}(2)_3, \tag{2.17}$$

and, correspondingly, we compute the new "bi-local" purity $P_2(|\psi\rangle)$ by using an appropriate basis of Ω_2 – including six bilocal Pauli operators in addition to the nine operators in Ω_1. How does the above classification change? Were qubit 3 not present, then clearly any pure state in $\mathcal{H}_1 \otimes \mathcal{H}_2$ would be disentangled relative to the full algebra $\mathfrak{su}(4)_{12}$. With qubit 3 included, one still expects any internal structure within qubits 1 and 2 to be irrelevant, as long as the pair reduced state remains pure. This intuition is reflected by the behavior of $P_2(|\psi\rangle)$, which now attains its maximum on both product states as before *and* the bi-separable class B_{12} – which thus becomes extremal. Both B_{23} and B_{13} become maximally generalized entangled relative to this observer, together with $|\mathtt{GHZ}\rangle$ ($P_2 = 1/3$), whereas the $|\mathtt{W}\rangle$ class shows intermediate GE with $P_2(|\mathtt{W}\rangle) = 11/27$.

Other interesting scenarios may be conceived, which resemble in part ordinary entanglement – in the sense that partial accessibility of subsystems in $\mathcal{H}$ is retained – yet differ in the identification of actual "local" degrees of freedom – in that observables spaces on different factors overlap. Suppose, for instance, that qubits 1, 2, 3 are spatially separated and arranged on a line, or reside at the vertexes of a triangle, with physical transformations generated by nearest-neighbor two-body couplings in the first case, or arbitrary two-body couplings in the second one. Then the resulting operational constraints are naturally captured by distinguished observable sets of the form

$$\Omega_3 = \mathrm{span}_{\mathbb{R}}\{\sigma_a^{(1)} \otimes \sigma_b^{(2)} \otimes \mathbb{1}^{(3)},\, \mathbb{1}^{(1)} \otimes \sigma_b^{(2)} \otimes \sigma_c^{(3)}\}, \tag{2.18}$$

or, respectively,

$$\Omega_4 = \mathrm{span}_{\mathbb{R}}\{\sigma_a^{(1)} \otimes \sigma_b^{(2)} \otimes \mathbb{1}^{(3)},\, \mathbb{1}^{(1)} \otimes \sigma_b^{(2)} \otimes \sigma_c^{(3)},\; \sigma_a^{(1)} \otimes \mathbb{1}^{(2)} \otimes \sigma_c^{(3)}\}. \tag{2.19}$$

A formal analysis of GE relative to the above distinguished sets is technically more difficult because neither of $\Omega_{3,4}$ is a Lie algebra. While such an analysis is beyond our current purposes, following the GE behavior of different classes of pure states as observables are added through the progression $\Omega_1 \to \Omega_2 \to \Omega_3 \to \Omega_4$ clearly serves to illustrate how GE is directly observer-dependent irrespective of whether a direct correspondence between the concept of "locality" and "locally accessible" subsystem degrees of freedom is possible: note, in particular, that relative to Ω_4 every bi-separable pure state of the three qubits becomes extremal – thus generalized unentangled, consistently with physical intuition.

2.4.1.3 The convex cones setting; GE in Popescu–Rohrlich boxes

Although the convex-cones setting is more general than quantum mechanics, in this setting one may still define a natural notion of one system's being a subsystem of another, and of a system's being a tensor product of two systems. In this case, however, the notion of tensor product is not unique. For our purposes, we will (as in [30]) allow as a tensor product of cones $C_1 \subset V_2$ and $C_2 \subset V_2$ *any* cone Γ in $V_1 \otimes V_2$, with base $\hat{\Gamma}$ containing the *separable* tensor product $\hat{C}_1 \otimes_{\mathrm{sep}} \hat{C}_2$, and contained in the *maximal* tensor product $\hat{C}_1 \otimes_{\max} \hat{C}_2$. Both the separable and the maximal tensor-product constructions admit a simple interpretation. The separable tensor product is the convex hull of the products $x \otimes y$, $x \in \hat{C}_2$, $y \in \hat{C}_2$. These product states are just the natural generalization of classical product distributions and quantum product states: for any pair of observables on C_1 and C_2, the joint probability given by such states, for outcome-pairs, factorizes as a product of marginal distributions for the two systems. The maximal tensor product is just $(C_1^* \otimes_{\mathrm{sep}} C_2^*)^*$, i.e., it is the set of unnormalized states that are non-negative on all "product outcomes." States in the maximal tensor product are precisely those that

do not allow signaling. This general notion of a tensor product appeared in [31], and (in a slightly different but essentially equivalent formalism of test spaces) the no-signaling construction of the maximal tensor product is done in [32, 33]. A review of the latter (with additional results on certain tensor products of two *quantum* systems) can be found in [34], together with many references to related work, and the notion of "system combination" independently developed in [35] should also be noted as closely related.

One can specialize our GE notion (and, correspondingly, generalize the notion of bipartite quantum entanglement) to such a tensor product, by letting $D \subset V_1 \otimes V_2$ be the cone generated by restricting states in Γ to the space of effects spanned by $\{x \otimes \lambda, \lambda \otimes y | x \in C_1^*, y \in C_2^*\}$. There is a natural map π satisfying the conditions stated in the definition of GE. For any state ω in Γ, its marginal states ω_1, ω_2 on system 1 or 2 may be defined as the restrictions of π to the spaces of observables on system 1 or 2, respectively, i.e., to $\{x \otimes \lambda | x \in C_1\}$ or $\{\lambda \otimes y | y \in C_2^*\}$. The reduced states $\pi(\omega) \in D$ are thus just pairs (ω_1, ω_2) of marginal states, and *whatever* tensor product between the maximal and the separable is chosen, it turns out that, relative to D and π, the unentangled bipartite states in that tensor product are just the ones in the *separable* tensor product. That is, pure unentangled states are just products of pure (extremal) states, and general unentangled states are convex combinations of these. This follows from the fact (see, e.g., Lemma 3 of [30], though this is almost certainly not its first appearance) that, if either marginal state of some state in a bipartite tensor product, as we have defined it, is pure, the bipartite state is a product state.

As a concrete non-quantum, non-classical example of the distinction between product and entangled pure states, consider the extremal states of a pair of two-output, two-input *Popescu–Rohrlich (PR) boxes*. In terms of convex sets, a single such box may be viewed as having a state space that is polyhedral with square base. This can be visualized in $\mathbb{R}^3$, although it is in some ways more natural to view it as a non-generating cone in $\mathbb{R}^4$, as the interpretation below will clarify. Suppose we have chosen an orthonormal basis, and let the distinguished $(\lambda(x) = 1)$ square base lie in the plane perpendicular to a line between the center of the square base and the origin. The cone of unnormalized measurement outcomes (the dual cone to this) is thus also a square polyhedral cone: if we visualize it in the same space and represent evaluation of functionals by the Euclidean inner product in this space, it will be rotated by $\pi/4$, relative to the primal cone, around the line through the center of the square and the origin (and also uniformly dilated or contracted relative to it). No matter how we affinely embed the primal cone in $\mathbb{R}^3$, it will not coincide with the dual cone determined by a Euclidean inner product, so this cone does not enjoy the important property of *self-duality*, but it is still isomorphic to its dual.

The interpretation of the states is in terms of two alternative measurements, each having two possible outcomes, the measurements labeled in the usual PR-box

formalism by an "input" bit specifying which measurement is to be done, and the outcome of each measurement by an "output" bit – so that, taken together, these two measurements have a total of four possible measurement outcomes, each corresponding to a pair of an input and an output bit. Pictorially, such outcomes correspond to points on the four extremal rays of the dual cones. View the square state space as embedded in $\mathbb{R}^2$, so that the vertices are $(0, 0)$, $(0, 1)$, $(1, 1)$, $(1, 0)$ as we go clockwise around the square base, and arbitrary states correspond to points $(p_0, p_1) \in [0, 1] \times [0, 1]$. Interpret p_0 as the probability of measurement 0 yielding the result 0, $(1 - p_0)$ as the probability of it yielding the result 1, p_1 as the probability of measurement 1 yielding 0, and $(1 - p_1)$ as the probability of measurement 1 yielding 1. Thus, the two-parameter representation of the state can be viewed as the result of an affine embedding of the subset of the four-parameter states $(p_{0|0}, p_{1|0}, p_{0|1}, p_{1|1})$, satisfying the normalization constraints $p_{0|0} + p_{1|0} = 1$, $p_{0|1} + p_{1|1} = 1$, into $\mathbb{R}^2$ (with $p_{i|j}$ denoting conditional probability). Similarly the cone of unnormalized states in $\mathbb{R}^3$ can be viewed as an isomorphic affine image (embedding) of the three-dimensional subspace of the four parameter states in the octant $\mathbb{R}^4_+$ satisfying the constraint $p_{0|0} + p_{1|0} = p_{0|1} + p_{1|1}$.

By a similar construction, analogous boxes can be defined for an arbitrary finite number of "inputs" N (corresponding to alternative measurements), and numbers $M_i, i \in \{1, \ldots, N\}$ of possible outputs for each measurement (usually taken to be a number M independent of i). Such sets of alternative measurements are also known in the quantum-logic literature as *semi-classical test spaces* [36].

The state space of a pair of such PR boxes is just the maximal tensor product of the state spaces of two PR boxes: it was introduced [37] in the PR-box literature as the set of all possible states of correlations between the outcomes of two PR boxes that do not permit signaling. This state space is a polytope; it has been studied in [38] and, in particular, its extremal points for the two-input, d-output case have been explicitly described. For our case, $d = 2$, there are two types of extremal states. They are best understood by exhibiting states as 4×4 matrices of probabilities $p_{ij|kl}$, where ij is the pair of Alice's outcome label, Bob's outcome label ("outputs"), and kl is the pair of Alice's measurement, Bob's measurement ("inputs"). Thus the matrix has a natural 2×2 block structure, the ijth block being the 2×2 matrix for the probabilities of the four possible Alice/Bob outcome-pairs when Alice does measurement k, Bob measurement l. That is, explicitly,

$$\left(\begin{array}{cc|cc} p_{00|00} & p_{01|00} & p_{00|01} & p_{01|01} \\ p_{10|00} & p_{11|00} & p_{10|01} & p_{11|01} \\ \hline p_{00|10} & p_{01|10} & p_{00|11} & p_{01|11} \\ p_{10|10} & p_{11|10} & p_{10|11} & p_{11|11} \end{array}\right). \tag{2.20}$$

The standard normalization condition is that the probabilities within each block sum to 1. Bob's marginal state is the vector of column sums of the upper half of the

matrix, interpreted as $(p_{0|0}, p_{1|0}, p_{0|1}, p_{1|1})$; the condition that Alice cannot signal to Bob translates into the fact that this is equal to the vector of column sums of the lower half of the matrix (i.e., Bob's probabilities are independent of Alice's choice of measurement). Similarly, Alice's marginals are the row sums of the left half of the matrix, equal by virtue of the no-signaling condition to those of the right half. The fact that these are linear equality constraints, together with the positivity constraint on probabilities, is what makes the no-signaling state space a polytope and, consequently, makes the cone of unnormalized states a polyhedral cone.

The extremal points were found and classified in [38] and, as mentioned, are of two types. One is represented by the following state:

$$\left(\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right). \tag{2.21}$$

Alice's reduced state is just the marginal state $(1, 0, 1, 0)$, as is Bob's; these are both extremal states, so this is generalized unentangled by our definition: indeed, it can also easily be seen to be a product state, $p_{ij|kl} = p^{\mathrm{A}}_{i|k} p^{\mathrm{B}}_{j|l}$. There are 16 representatives of this class, representing products of each of the four local extremal states for Alice with each of four local extremal states for Bob.

The other class consists of eight entangled extremal states. These are all locally equivalent to the representative[4]

$$\left(\begin{array}{cc|cc} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \hline \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{array}\right). \tag{2.22}$$

In this case, both Alice's and Bob's marginals are the mixed state $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, showing (given that these are extremal overall, as proved in [38]) that this state is generalized entangled by our definition. Of course, the fact that this is entangled in the sense that it is a pure non-product state was already observed in [38]. As noted there, the other 15 pure product states, and the other 7 pure entangled states, can be obtained from the above two states by local transformations consisting of relabeling the measurements and the outcomes. The entangled states, for example, may be described via the shorthand A (for anticorrelated) for the 2×2 matrix that looks like $\sigma_x/2$, and C (for correlated) for the 2×2 matrix $\mathbb{1}/2$. In this way, the above entangled state reads:

$$\left(\begin{array}{cc} C & C \\ C & A \end{array}\right), \tag{2.23}$$

[4] Note that the terminology is not uniform; in particular, extremal entangled states as defined in (2.22) are commonly referred to as the "PR box."

and all eight entangled states may be obtained by noting that relabeling the measurements just interchanges rows and/or columns – allowing us, with a few such relabellings, to place the A in any of the four places, whereas relabeling the outcomes of one of Alice's or Bob's measurements interchanges C and A in the row (for Alice) or column (for Bob) corresponding to that measurement – allowing us access to the four states having three As and one C.

Reference [38] also describes the single local equivalence class of entangled pure states of a pair of two-measurement, d-outcome boxes, and some of the 44 local equivalence classes of entangled pure states of three two-measurement, two-outcome boxes, together with many other interesting results, for example on local interconversion of various combinations of states of PR boxes. Further properties of operational theories based on PR boxes are investigated in [35].

2.4.2 Entanglement without locality ...

The extent to which the GE notion genuinely enlarges the standard subsystem-based framework may be further appreciated in situations where $\mathcal{H}$ is intrinsically indecomposable – thus no factorization of $\mathcal{H}$ exists and conventional entanglement is not directly applicable. A physically motivated example is offered by a quantum spin-J system (take for definiteness $J \in \mathbb{N}$), whose $d = (2J+1)$-dimensional state space carries an irreducible representation of $\mathfrak{su}(2)$, with angular-momentum generators J_α satisfying commutation rules $\left[J_\alpha, J_\beta\right] = 2\mathrm{i}\epsilon_{\alpha\beta\gamma}J_\gamma$, $\epsilon_{\alpha\beta\gamma}$ denoting the completely antisymmetric tensor. Because the GE notion rests only on convex properties of sets of quantum states and observables, the definition of GE is still applicable as soon as distinguished observable sets are specified. In particular, the full algebra of traceless observables $\mathfrak{g}$ is isomorphic to $\mathfrak{g} \simeq \mathfrak{su}(d)$: clearly, expectations in $\mathfrak{g}$ completely determine an arbitrary pure state in $\mathcal{H}$ – accordingly, every pure state is (extremal) unentangled relative to $\mathfrak{g}$.

Imagine, however, that only expectations of observables in the "local" subalgebra $\mathfrak{h} = \mathfrak{su}(2)$ are available. In this case, $\mathfrak{h}$-reduced states may be identified with three-dimensional vectors of expectation values $\langle J_\alpha \rangle$, $\alpha = x, y, z$, which form a ball of radius J in $\mathbb{R}^3$. Using the notation $\mathbf{J}$ for the vector (J_x, J_y, J_z), these are vectors $\langle \mathbf{J} \rangle$. In addition, $\mathbf{J}^2 = \mathbf{J} \cdot \mathbf{J} = J_x^2 + J_y^2 + J_z^2$. The extremal states correspond geometrically to points on the surface of such a ball, that is to the GCSs of SU(2) – also called *spin coherent states* (SCSs) [27] – which are characterized by maximal spin projection along a given component,

$$(\mathbf{n} \cdot \mathbf{J})|\psi\rangle_{\mathrm{SCS}} = J|\psi\rangle_{\mathrm{SCS}}, \quad |\mathbf{n}| = 1.$$

For any given choice of spin direction, e.g., $\mathbf{n} = \hat{z}$, let $\{|J, J\rangle, |J, J-1\rangle, \ldots, |J, 0\rangle, \ldots, |J, -J+1\rangle, |J, -J\rangle\}$ be an orthonormal basis of $\mathcal{H}$ consisting of

joint $\mathbf{J}^2$, J_z eigenstates. Then states $|J, \pm J\rangle$ lie on the surface of the ball and therefore are generalized unentangled, whereas the remaining states lie on the inside, and hence are not extremal: in particular, the state $|J, 0\rangle$ lies at the middle of the ball and is *maximally entangled* relative to $\mathfrak{su}(2)$. Mathematically, this is reflected in minimal $\mathfrak{su}(2)$-purity,

$$P_{\mathfrak{su}(2)}(|J, 0\rangle) = \frac{1}{J^2} \sum_{\alpha=x,y,z} \langle J, 0|J_\alpha|J, 0\rangle^2 = 0.$$

Physically, the presence of GE captures the fact that no "local" unitary operation (no rotation in SU(2)) is able to connect such a state with extremal states on the surface: achieving that requires "entangling" operations generated by Hamiltonians in the full $\mathfrak{g}$.

Note that, because $\mathfrak{h}$ is irreducible, the $\mathfrak{h}$-purity coincides, up to an additive constant, with the quantity

$$(\Delta\mathcal{I})^2 = \sum_\alpha \left[\langle J_\alpha^2\rangle - \langle J_\alpha\rangle^2\right] = J(J+1) - J^2 P_{\mathfrak{su}(2)},$$

which measures the so-called *invariant uncertainty* of SU(2) and is minimized by GCSs [27, 39]. From this point of view, generalized unentangled (maximally entangled) states may thus be seen as maximally close to (remote from) "classical reality" – as measured in terms of the corresponding minimal (maximal) amount of quantum fluctuations [18, 19].

2.4.3 Separability without entanglement ...

A further element of distinction between conventional and generalized entanglement arises in situations where a local structure exists in $\mathcal{H}$ as in (2.13), but operational capabilities are restricted to a *proper subset of local observables*.

Consider, for instance, two spin-J particles, with associated bipartite Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$, $d_0 = 2J + 1$, and full observable Lie algebra $\mathfrak{g} \simeq \mathfrak{su}(d^2)$. Arbitrary (subsystem)-local observables correspond to $\mathfrak{h}_{\rm loc} \simeq \mathfrak{su}(d)_1 \oplus \mathfrak{su}(d)_2$. If, as in the above example, only observables linear in the basic angular-momentum generators are distinguished for each subsystem, then the accessible "local" sub-algebra is

$$\mathfrak{h}'_{\rm loc} = \mathfrak{su}(2)_1 \oplus \mathfrak{su}(2)_2 \subset \mathfrak{h}_{\rm loc}.$$

The set of generalized unentangled states relative to $\mathfrak{h}'_{\rm loc}$ consists of states of the form $|\mathsf{SCS}\rangle_1 \otimes |\mathsf{SCS}\rangle_2$. However, other *tensor-product* states like $|J, 0\rangle_1 \otimes |J, 0\rangle_2$ are maximally entangled relative to the local sub-algebra: to fully "resolve" the product nature of such states requires access to expectations of arbitrary observables in $\mathfrak{h}_{\rm loc}$.

2.4.4 Fermionic entanglement

As a last example, we reconsider the simplest instance of entanglement in systems of indistinguishable fermions, which was briefly mentioned at the end of Section 2.2.

Because the requirement of complete antisymmetry of the joint state vector under particle exchange is automatically incorporated, identical (spinless) fermions are most naturally described in a second-quantized language based on canonical fermion operators. Consider two identical fermions, each of which may occupy one of two modes [24]. For each mode $j = 1, 2$, let $c_j^\dagger$ and c_j denote creation and annihilation operators, respectively, obeying the following anticommutation rules:

$$\begin{aligned} \left[c_i, c_j\right]_+ &= \left[c_i^\dagger, c_j^\dagger\right]_+ = 0, \\ \left[c_i^\dagger, c_j\right]_+ &= \delta_{ij}, \end{aligned}$$

where $[\,,\,]_+$ denotes the anticommutator. Let in addition $|\mathsf{vac}\rangle$ denote a reference "vacuum" state containing no fermions. For instance, we may associate mode 1 with Alice and mode 2 with Bob, who reside at spatially separated locations. Then a state where a fermion is created at Alice's site corresponds to $|c_1^\dagger\rangle \equiv c_1^\dagger|\mathsf{vac}\rangle$, and so on.

What kind of entanglement is physically meaningful in fermionic states? Insight into this question may be sought by changing the description into a more familiar spin language, by exploiting the Jordan–Wigner isomorphic mapping between the above fermionic operators and the Pauli algebra,

$$\begin{aligned} S_+^{(1)} &= c_1^\dagger, & S_-^{(1)} &= (S_+^{(1)})^\dagger, \\ S_+^{(2)} &= (1 - 2n_1)c_2^\dagger, & S_-^{(2)} &= (S_+^{(2)})^\dagger, \end{aligned}$$

where $\hat{n}_1 = c_1^\dagger c_1$ is the fermion number operator for mode 1, and the operators $S_\pm^{(j)} = (\sigma_x^{(j)} \pm \mathrm{i}\sigma_y^{(j)})/2$ so constructed obey $\mathfrak{su}(2)$ commutation rules. Then the following correspondence may be established between states in the usual spin-$\frac{1}{2}$ basis $\{|0\rangle, |1\rangle\}$ and fermionic states:

$$\begin{aligned} |00\rangle &\leftrightarrow |\mathsf{vac}\rangle, \\ |01\rangle &\leftrightarrow c_1^\dagger|\mathsf{vac}\rangle, \\ |10\rangle &\leftrightarrow c_2^\dagger|\mathsf{vac}\rangle, \\ |11\rangle &\leftrightarrow c_1^\dagger c_2^\dagger|\mathsf{vac}\rangle. \end{aligned} \tag{2.24}$$

Presumably, we would want none of these states (containing, respectively, zero, one, one, and two fermions at different locations) to be "entangled" by any

reasonable definition. What about the set of Bell states given in (2.3)? By applying the Jordan–Wigner mapping, these can be rewritten as follows:

$$|\Phi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle) \leftrightarrow \frac{1}{\sqrt{2}}(c_1^\dagger \pm c_2^\dagger)|\mathsf{vac}\rangle,$$
$$|\Psi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle) \leftrightarrow \frac{1}{\sqrt{2}}(\mathbb{1} \pm c_1^\dagger c_2^\dagger)|\mathsf{vac}\rangle = \frac{1}{\sqrt{2}}\left(|\mathsf{vac}\rangle \pm c_1^\dagger c_2^\dagger|\mathsf{vac}\rangle\right).$$

Note that states in the upper line are still states with a *fixed* number of particles (one particle), which are described in general by so-called "Slater determinants" in the condensed-matter terminology. States in the bottom line, in contrast, are linear combinations of states with zero and two fermions that is, linear combinations of Slater determinants with *different* particle number.

Suppose we consider an arbitrary physical scenario or process in which *no* change of the fermion number in a state can occur. Then the admissible physical observables for fermions must commute with the total fermion operator,

$$\hat{N} = \hat{n}_1 + \hat{n}_2 \to \frac{\sigma_z^{(1)}}{2} + \frac{\sigma_z^{(2)}}{2} - \mathbb{1} = S_z - \mathbb{1},$$

which recovers precisely the conservation law discussed in Section 2.2. In fact, bilinear fermion operators of the form $\{c_i^\dagger c_j, 1 \le i, j \le 2\}$ may be seen to satisfy $\mathfrak{u}(2)$-commutation rules [23], which makes the latter a natural candidate for a distinguished fermionic observable set. A larger fermionic algebra arises if arbitrary bilinear fermion operators are included (e.g., operators of the form $c_i^\dagger c_j^\dagger$), leading to $\mathfrak{so}(4) \supset \mathfrak{u}(2)$. Written in the fermionic language, the $\mathfrak{u}(2)$ Lie algebra given in (2.10) becomes

$$\mathfrak{u}(2) = \mathrm{span}_{\mathbb{R}}\{\hat{n}_1 - 1/2, \hat{n}_2 - 1/2, (c_1^\dagger c_2 + c_2^\dagger c_1)/\sqrt{2}, \ \mathrm{i}(c_1^\dagger c_2 - c_2^\dagger c_1)/\sqrt{2}\}.$$

The corresponding $\mathfrak{u}(2)$-purity, $P_{\mathfrak{u}(2)}$ (computed either in the fermionic or in the spin language), attains its maximum on both the number eigenstates of (2.24) and the two Bell states $|\Phi^{\pm}\rangle$. Thus, the latter states are certainly "mode-entangled" with respect to the local spin algebra Ω_{loc}, but generalized unentangled relative to the fermionic algebra, which is consistent with the original intuition about their one-particle nature. In contrast, states $|\Psi^{\pm}\rangle$ have zero $\mathfrak{u}(2)$-purity; thus they are maximally entangled both in the conventional spin sense *and relative to the* $\mathfrak{u}(2)$ *observer*. That is, states $|\Psi^{\pm}\rangle$ contain genuine fermionic entanglement – irrespective of the operator language used. Physically, this expresses the fact that a linear combination of Slater determinants with different fermion numbers *cannot* be distinguished from a mixture by relying solely on expectations in $\mathfrak{u}(2)$: indeed, distinguishing would involve expectations of operators (like, e.g., $\sigma_x^{(1)} \otimes \sigma_x^{(2)}$ in $\mathfrak{so}(4)$)

that have non-zero matrix elements between states of different fermion number, hence requiring the above conservation law to be broken.[5]

2.5 Generalized entanglement: applications and implications (so far . . .)

The GE notion has proved so far a powerful unifying framework for linking entanglement properties to various physical, information-theoretic, and conceptual aspects of "complexity" and "classicality" emerging in a variety of scenarios.

2.5.1 Complexity implications

Following one of the original motivations underlying GE [23, 24], purity measures associated to appropriate observable subspaces are providing novel diagnostic tools for characterizing the "correlations" present in eigenstates of interacting quantum many-body Hamiltonians – allowing natural contact to be established with state-complexity notions, such as the *number of principal components* and the *inverse participation ratio*, borrowed from quantum statistical physics [40]. Two situations are especially relevant in this regard, and have been the subject of intense investigation recently (see, e.g., [41–44] and references therein). On the one hand, the correlations present in the *ground state* of a many-body system may undergo a structural change as some parameters in the Hamiltonian are changed at zero temperature across "critical" values – giving rise to so-called "quantum phase transitions" in the limit of infinite system size. In this case, natural Lie-algebraic GE measures constructed from fermionic and/or spin operators have made it possible to successfully identify and characterize the ensuing critical behavior in a large class of transitions induced by a spontaneous symmetry breaking [23].[6] On the other hand, structural changes may also take place for *any* typical many-body state if the addition of a perturbation or disorder to the original Hamiltonian causes a crossover from a regular, "integrable" regime into non-integrability and so-called "quantum chaos." The onset of chaos may in turn manifest itself at both the static level – in terms, for instance, of different eigenvector statistics as described by so-called "random matrix theory" [45] – and at the dynamic level – in terms of different behavior of quantum "fidelity" as a function of time [46–48], or hypersensitivity to

[5] Notice that operators commuting with $S_z \leftrightarrow \hat{n}$ suffice for distinguishing states $|\Phi^{\pm}\rangle$ from mixtures. The importance of taking into account the constraint of fermion-number conservation toward consistently defining entanglement in states of identical fermions has been stressed independently by Noah Linden (private communication).

[6] For the ground states of a paradigmatic class of one-dimensional critical spin-$\frac{1}{2}$ chains with N sites (the so-called XY model in a transverse field), the N-sites fermionic purity $P_{\mathrm{u}(N)}$ may be directly related to the *variance* of the total fermion-number operator $\hat{N}$ [23]. Thus, the relevant fermionic entanglement in this case directly originates from the superposition of states with different fermion number.

perturbations [49, 50]. Investigations in paradigmatic systems such as disordered quantum spin lattices [51] and chaotic quantum maps [52] have given clear indications so far that GE and GE generation with respect to appropriate observables can serve as reliable indicators for detecting and probing quantum chaos.

Suggestively, the GE notion has also recently shed light on the conditions needed to unlock the full power of quantum computational models as compared with purely classical ones. For a large class of "Lie-algebraic quantum computations," which are specified through controllable interactions and measurable observables in a Lie algebra $\mathfrak{g}$ and initialization of the system in a GCS of $\mathfrak{g}$, the results in [53] demonstrate how GE is required in order for genuinely stronger-than-classical computational models to emerge.

2.5.2 Classicality implications

If we consider a quantum system with the distinguished set of observables consisting of all those observables diagonal in some *fixed* basis (a special case of the associative-algebraic setting, in fact a case where the associative algebra is commutative), the generalized unentangled states are precisely the set of density matrices diagonal in that basis. Often the selection of such a distinguished basis is viewed as the selection of a distinguished "classical" set of states because the dynamics that preserve diagonality of density matrices in such a basis, together with measurements of observables diagonal in the same basis, are equivalent to a classical theory on a number of classical states equal to the dimension. This notion also generalizes to the convex-cones setting as follows. In [30], clonable sets of states were shown to be jointly distinguishable by a single measurement, and vice versa. Also, broadcastable sets of states were shown to belong [30] to the convex hull of such a set of jointly distinguishable states (which we term a "simplex generated by distinguishable states" (SGDS)). Indeed, for any map B from a convex set Ω to a tensor product $\Omega \otimes \Omega$, the set of states broadcast by B was shown to be *precisely* such a simplex (though for some maps it will be the empty set, viewed as a degenerate case of such a simplex). A (finitely generated) simplex is the convex hull of n or fewer points in $\mathbb{R}^n$, for some finite n; the theory whose set of normalized states is an n-simplex in $\mathbb{R}^n$ and whose cone of measurements is the dual of this (which is also based on an n-simplex) is classical. Because the projectors $|i\rangle\langle i|$ onto the states of some orthonormal basis form a d-dimensional simplex in the d^2-dimensional space of Hermitian operators on a d-dimensional quantum system, and because they are distinguishable by a single measurement, they are a special case of an SGDS as defined above. For an SGDS generated by states $\omega_i \in C$, if $a_i \in C^*$ such that $\sum_i a_i = \lambda$ is a measurement distinguishing the states ω_i, then we may take as our cone D the cone of reduced states, defined as the restrictions of states in C to the

space of functionals in the span of the a_i. Then states in the SGDS are precisely the generalized unentangled states relative to the observables generated by the a_i: that is, classical states with respect to this set of distinguished observables.

The problem of identifying classicality aspects of generalized unentangled states has also been tackled recently by making explicit contact with the open quantum system theory and the decoherence program [54]. Within a Lie-algebraic formulation of Markovian quantum dynamics, in particular, GCSs are found to minimize an invariant uncertainty that is closely related to their quantum Fisher-information content, and they are seen to emerge as most predictable "einselected" pointer states under appropriate conditions on the interaction between the system and its surrounding environment – generalizing the characterization of canonical (harmonic-oscillator) coherent states as most stable (hence most classical) states in the presence of decoherence [39].

2.5.3 Conceptual implications and unresolved problems

While all the above characterizations consistently tie the absence of GE to aspects of classicality as different as minimum state complexity, classical simulatability, minimum uncertainty, and maximum predictability, an independent notion of classicality is, in principle, failure to violate appropriate *inequalities* violated by quantum mechanics: for *pure* states, a Bell-type inequality is violated if and only if the state is not separable in the conventional sense [55, 56]. Assessing whether the presence of GE relative to appropriate observables may also be linked to the violation of suitable "generalized Bell inequalities" is, from both a fundamental and a philosophical perspective, one of the main open questions about GE at present (see also [18] for some ideas along these lines). From the point of view of QIS applications, obtaining a more thorough resource-based characterization of GE, linking the presence of GE to tasks that would not be achievable otherwise, and/or quantifying the amount and type of GE required to accomplish them constitute important areas for exploration where additional fundamental insight about GE is expected to emerge. In particular, a resource-based approach may both further elucidate the meaning of GE in a single-quantum system [19, 52] and shed light on ways for distinguishing (conceptually and operationally) between genuinely "quantum" and "classical" correlation aspects (see, e.g., [57–59]) within GE.

As we stressed throughout this work, the GE approach is naturally suited to defining entanglement in settings more general than standard quantum mechanics – in particular, abstract operational theories based on convex structures. It may be worth observing that, by abandoning objective, absolute notions of properties such as locality and separability – by acknowledging instead that different observers can give different descriptions of these concepts and the ensuing notion

of entanglement, the GE approach also shares some of its motivations with the recently proposed approach of *relational quantum mechanics* [60, 61]. While important differences between the latter and operational theories exist in terms of how observers themselves are included and described, it could be intriguing to further scrutinize differences and similarities between the two approaches in the light of the GE concept. Additional open questions and implications for quantum foundations are discussed in [14, 21].

2.6 Conclusion

We believe that the richness of applications as well as the numerous questions raised by the GE program to date speak for themselves about the significance and potential of GE for properly capturing the unavoidable relativity of entanglement. We hope that the GE approach will provide fresh stimulus for the exploration of entanglement to be extended into still-unexplored physical, mathematical, and philosophical regions.

Acknowledgments

It is a pleasure to thank Manny Knill, Gerardo Ortiz, and Rolando Somma for collaboration and interaction on various aspects of the GE program. LV is also indebted to Abner Shimony, Noah Linden, and Sandu Popescu for thought-provoking discussions and exchanges. Work at Los Alamos is supported in part through the Laboratory Directed Research and Development program.

References

[1] A. Einstein, B. Podolsky, and N. Rosen, Can quantum-mechanical description of physical reality be considered complete?, *Phys. Rev.* **47**, 777 (1935).
[2] E. Schrödinger, Die gegenwärtige Situation in der Quantenmechanik, *Naturwissenschaften* **23**, 807 (1935).
[3] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge: Cambridge University Press, 1993).
[4] J. F. Clauser and A. Shimony, Bell's theorem: experimental tests and implications, *Rep. Prog. Phys.* **41**, 1881 (1978).
[5] A. Aspect, P. Grangier, and G. Roger, Experimental realization of Einstein–Podolsky–Rosen–Bohm *Gedankenexperiment*: a new violation of Bell's inequalities, *Phys. Rev. Lett.* **47**, 460 (1981).
[6] S. Popescu and D. Rohrlich, The joy of entanglement, in *Introduction to Quantum Computation and Information*, H.-K. Lo, S. Popescu, and T. Spiller (eds.) (Singapore: World Scientific, 1998).
[7] C. H. Bennett and S. Wiesner, Communication via one- and two-particle operators on EPR states, *Phys. Rev. Lett.* **69**, 2881 (1992).

[8] C. H. Bennett, G. Brassard, C. Crépeau *et al.*, Teleporting an unknown quantum state via dual classical and EPR channels, *Phys. Rev. Lett.* **70**, 1895 (1993).
[9] R. Jozsa and N. Linden, On the role of entanglement in quantum computational speed-up, *Proc. Roy. Soc. London A* **454**, 2011 (2003).
[10] K. Eckert, J. Schliemann, D. Bruss, and M. Lewenstein, Quantum correlations in systems of indistinguishable particles, *Ann. Phys.* **299**, 88 (2002) and references therein.
[11] P. Zanardi, Quantum entanglement in fermionic lattices, *Phys. Rev. A* **65**, 042101 (2002).
[12] H. Wiseman and J. A. Vaccaro, Entanglement of indistinguishable particles shared between two parties, *Phys. Rev. Lett.* **91**, 097902 (2003).
[13] M. Kindermann, Proposal of an experimentally accessible measure of many-fermion entanglement, *Phys. Rev. Lett.* **96**, 240403 (2006).
[14] H. Barnum, E. Knill, G. Ortiz, R. Somma, and L. Viola, Generalizations of entanglement based on coherent states and convex sets, *Phys. Rev. A* **68**, 032308 (2003).
[15] H. Barnum, E. Knill, G. Ortiz, R. Somma, and L. Viola, A subsystem-independent generalization of entanglement, *Phys. Rev. Lett.* **92**, 107902 (2004).
[16] L. Viola, E. Knill, and R. Laflamme, Constructing qubits in physical systems, *J. Math. Phys.* **34**, 7067 (2001).
[17] P. Zanardi, D. A. Lidar, and S. Lloyd, Quantum tensor product structures are observable-induced, *Phys. Rev. Lett.* **92**, 060402 (2004).
[18] A. Klyachko, Coherent states, entanglement, and geometric invariant theory, arXiv.org e-print quant-ph/0206012.
[19] A. S. Shumovsky, Observables, quantum fluctuations, and "quantum reality", *Conc. Phys.* **1**, 151 (2004); A. A. Klyachko and A. S. Shumovsky, General entanglement, *J. Phys.: Conf. Series* **36**, 87 (2006) and references therein.
[20] H. Barnum, G. Ortiz, R. Somma, and L. Viola, A generalization of entanglement to convex operational theories: entanglement relative to a subspace of observables, *Int. J. Theor. Phys.* **44**, 2127 (2005).
[21] H. Barnum, The view from everywhere: convex operational theories, quantum information, quantum foundations, and the coordination of quantum agents' perspectives, in *Proceedings of the International Conference "Quantum Theory: Reconsideration of Foundations – 2"*, Yu. A. Khrennikov (ed.) (Växjö: Växjö University Press, 2004), pp. 553–637, arXiv.org e-print quant-ph/0611110.
[22] L. Viola, H. Barnum, E. Knill, G. Ortiz, and R. Somma, Entanglement beyond subsystems, *Contemp. Math.* **381**, 117 (2005).
[23] R. Somma, G. Ortiz, H. Barnum, E. Knill, and L. Viola, Nature and measure of entanglement in quantum phase transitions, *Phys. Rev. A* **70**, 042311 (2004).
[24] G. Ortiz, R. Somma, H. Barnum, E. Knill, and L. Viola, Entanglement as an observer-dependent concept: an application to quantum phase transitions, in *CMT27 (Toulouse) Workshop Proceedings* (New York: Nova Science Publisher, 2004).
[25] R. Somma, H. Barnum, E. Knill, G. Ortiz, and L. Viola, Generalized entanglement and quantum phase transitions, *Int. J. Mod. Phys. B* **20**, 2760 (2006).
[26] C. Batista and G. Ortiz, Generalized Jordan–Wigner transformations, *Phys. Rev. Lett.* **86**, 1802 (2001) and references therein.
[27] A. Perelomov, *Generalized Coherent States* (Berlin: Springer-Verlag, 1985).
[28] G. Brennen, An observable measure of entanglement for pure states of multi-qubit systems, *Quantum Inf. Comput.* **3**, 619 (2003).
[29] D. A. Meyer and N. R. Wallach, Global entanglement in multiparticle systems, *J. Math. Phys.* **43**, 4273 (2002).

[30] H. Barnum, J. Barrett, M. S. Leifer, and A. Wilce, Cloning and broadcasting in generic probabilistic theories, arXiv.org e-print quant-ph/0509211.
[31] I. Namioka and R. R. Phelps, Tensor products of compact convex sets, *Pacific J. Math.* **9**, 469 (1968).
[32] D. J. Foulis and C. H. Randall, Empirical logic and tensor products, in *Interpretations and Foundations of Quantum Theory*, B. Hartkamper and H. Neumann (eds.), (Mannheim: Bibliographisches Institut, 1980).
[33] C. H. Randall and D. J. Foulis, Operational statistics and tensor products, in *Interpretations and Foundations of Quantum Theory*, H. Neumann (ed.) (Mannheim: Bibliographisches Institut, 1981).
[34] H. Barnum, C. Fuchs, J. Renes, and A. Wilce, Influence-free states on compound quantum systems, arXiv.org e-print quant-ph/0507108, submitted to *Phys. Rev. A.*
[35] J. Barrett, Information processing in generalized probabilistic theories, *Phys. Rev. A* **75**, 032304 (2007).
[36] A. Wilce, Test spaces and orthoalgebras, in *Current Research in Operational Quantum Logic*, B. Coecke, D. Moore, and A. Wilce (eds.) (Dordrecht: Kluwer, 2000).
[37] S. Popescu and D. Rohrlich, Quantum non-locality as an axiom, *Foundations Phys.* **24**, 379 (1994).
[38] J. Barrett, N. Linden, S. Massar, *et al.*, Non-local correlations as an information-theoretic resource, *Phys. Rev. A* **71**, 022101 (2005).
[39] S. Boixo, L. Viola, and G. Ortiz, Generalized coherent states as preferred states of open quantum systems, *Europhys. Lett.* **79**, 40003 (2007).
[40] L. Viola and W. G. Brown, Generalized entanglement as a framework for complex quantum systems: purity vs delocalization measures, *J. Phys. A* **40**, 8109 (2007).
[41] T. J. Osborne and M. A. Nielsen, Entanglement in a simple quantum phase transition, *Phys. Rev. A* **66**, 032110 (2002).
[42] G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev, Entanglement in quantum critical phenomena, *Phys. Rev. Lett.* **90**, 227902 (2003).
[43] L. Amico, R. Fazio, A. Osterloh, and V. Vedral, Entanglement in many-body systems, *Rev. Mod. Phys.* **80**, 517 (2008).
[44] C. Mejia-Monasterio, G. Benenti, G. G. Carlo, and G. Casati, Entanglement across a transition to quantum chaos, *Phys. Rev. A* **71**, 062324 (2005).
[45] F. Haake, *Quantum Signatures of Chaos* (Berlin: Springer-Verlag, 2000).
[46] A. Peres, Instability of quantum motion of a chaotic system, in *Quantum Chaos*, H. A. Cerdeira, R. Ramaswamy, M. C. Gutzwiller, and G. Casati (eds.) (Singapore: World Scientific, 1991).
[47] F. M. Cucchietti, H. M. Pastawski and R. A. Jalabert, Universality of the Lyapunov regime for the Loschmidt echo, *Phys. Rev. E* **70**, 035311 (2005).
[48] T. Gorin, T. Prosen, T. H. Seligman, and M. Znidaric, Dynamics of Loschmidt echoes and fidelity decay, *Phys. Rep.* **435**, 33 (2006).
[49] R. Schack and C. M. Caves, Information-theoretic characterization of quantum chaos, *Phys. Rev. E* **53**, 3257 (1996).
[50] A. J. Scott, T. A. Brun, C. M. Caves and R. Schack, Hypersensitivity and chaos signatures in the quantum baker's maps, *J. Phys. A* **39**, 13405 (2006).
[51] S. Montangero and L. Viola, Multipartite entanglement generation and fidelity decay in disordered qubit systems, *Phys. Rev. A* **73**, (R)040302 (2006).
[52] Y. S. Weinstein and L. Viola, Generalized entanglement as a framework for exploring quantum chaos, *Europhys. Lett.* **76**, 746 (2006).

[53] R. Somma, H. Barnum, E. Knill, and G. Ortiz, Efficient solvability of Hamiltonians and limits on the power of some quantum computational models, *Phys. Rev. Lett.* **97**, 190501 (2006).
[54] W. H. Zurek, Decoherence, einselection, and the quantum origins of the classical, *Rev. Mod. Phys.* **75**, 715 (2003).
[55] S. Popescu and D. Rohrlich, Generic quantum non-locality, *Phys. Lett. A* **166**, 293 (1992).
[56] R. F. Werner and M. Wolf, Bell inequalities and entanglement, arXiv.org quant-ph/0107093
[57] H. Ollivier and W. H. Zurek, Quantum discord: a measure of the quantumness of correlations, *Phys. Rev. Lett.* **88**, 017901 (2001).
[58] B. Groisman, S. Popescu, and A. Winter, Quantum, classical, and total amount of correlations in a quantum state, *Phys. Rev. A* **2**, 032317 (2005).
[59] M. Horodecki, P. Horodecki, R. Horodecki *et al.*, Local versus non-local information in quantum-information theory: formalism and phenomena, *Phys. Rev. A* **71**, 062307 (2005).
[60] C. Rovelli, Relational quantum mechanics, *Int. J. Theor. Phys.* **35**, 1637 (1996).
[61] M. Smerlak and C. Rovelli, Relational EPR, *Found. Phys.* **37**, 427 (2007).

3

Formalism locality in quantum theory and quantum gravity

Lucien Hardy

3.1 Introduction

Indefinite causal structure poses particular problems for theory formulation since many of the core ideas used in the usual approaches to theory construction depend on having definite causal structure. For example, the notion of a state across space evolving in time requires that we have some definite causal structure so we can define a state on a space-like hypersurface. We will see that many of these problems are mitigated if we are able to formulate the theory in a *formalism-local* (or F-local) fashion. A formulation of a physical theory is said to be F-local if, in making predictions for any given arbitrary space-time region, we need only refer to mathematical objects pertaining to that region. This is a desirable property both on the grounds of efficiency and since, if we have indefinite causal structure, it is not clear how to select some other space-time region on which our calculations may depend. The usual ways of formulating physical theories (the time-evolving state picture, the histories approach, and the local-equations approach) are not F-local.

We set up a framework for probabilistic theories with indefinite causal structure. This, the causaloid framework, is F-local. We describe how quantum theory can be formulated in the causaloid framework (in an F-local fashion). This provides yet another formulation of quantum theory. This formulation, however, may be particularly relevant to the problem of finding a theory of quantum gravity. The problem of quantum gravity is to find a theory that reduces in appropriate limits to general relativity and quantum theory (including, at least, those situations in which those two theories have been experimentally confirmed). To be significant, the theory must also make correct predictions for new experiments in the future.

The problem of combining two less fundamental theories into a more fundamental one is not something for which a simple algorithm can exist and thus we need a motivating idea to get started. Here we note that general relativity and quantum theory are each conservative and radical in complementary ways. General relativity

Philosophy of Quantum Information and Entanglement, ed. A. Bokulich and G. Jaeger. Published by Cambridge University Press.

is conservative in that it is deterministic but radical in that it has non-fixed causal structure (whether a particular interval is time-like is not fixed in advance but can be decided only after we have solved for the metric). Quantum theory is conservative in that it has fixed causal structure built in, but radical in that it is inherently probabilistic (standard quantum theory cannot be formulated without reference to probabilities). It seems likely that a theory of quantum gravity must inherit the radical features of the two component theories. Hence, we are looking for a theoretical structure that

(1) is probabilistic and
(2) has non-fixed causal structure.

In fact, we expect the situation to be even more radical. In general relativity the causal structure is not fixed in advance, but, once determined, there is a definite answer to the question of whether an interval is time-like or not. However, in quantum theory any quantity that is subject to variation is also subject to quantum uncertainty. This means that, in a theory of quantum gravity, there may be no matter of fact as to whether a particular interval is time-like or not. It is likely that the causal structure is not only non-fixed, but also indefinite. The fact that we expect the conservative features of each component theory to be replaced by the radical features in the other suggests that a theory of quantum gravity cannot be entirely formulated within general relativity or quantum theory. In this, our program differs from string theory [1] and loop quantum gravity [2], where the attempt is to formulate quantum gravity within quantum theory (though there are other approaches that, to varying extents, do not assume that quantum theory will remain intact [3–6]).

One signature of the fixed causal structure in quantum theory is the fact that we have a fixed background time t used to evolve the state $|\psi(t)\rangle = U(t)|\psi(0)\rangle$. A deeper signature of fixed causal structure in quantum theory can be seen by considering the different ways in which operators can be put together. The operators corresponding to two space-like separated regions are combined with the tensor product $A \otimes B$. If a system passes through two immediately sequential time-like separated regions then the appropriate way to combine the corresponding operators is with the direct product CB. If a system passes through two time-like separated regions that have a gap in between (i.e., they are not immediately sequential) then the appropriate way to combine the operators is with what we will call the question-mark product $[D?B]$. This linear operator is defined by $[D?B]\,C \equiv DCB$. For each situation, we must combine the associated operators in a way that depends on the causal relationship between the two regions. It would be good to have a mathematical framework that treats each type of situation on an equal footing since then the fixed causal background need not be ingrained into the very structure of the theory.

The task becomes one of finding a theoretical framework for probabilistic theories with indefinite causal structure that correlate recorded data. The causaloid framework set up in [7] (see also [8]) does this. In this framework the *causaloid product* is defined. This unifies the three products mentioned above from quantum theory (in the context of a more general mathematical framework).

In this chapter I will discuss the challenges posed by having indefinite causal structure and show how the causaloid formalism deals with them. I will indicate how the quantum theory of pairwise interacting qubits can be dealt with in this formalism (this is an important example since we can use it to do universal quantum computation). Finally we will look at the road to formulating quantum gravity in this framework.

3.2 Dealing with indefinite causal structure

Indefinite causal structure is much more radical than merely having non-fixed causal structure as in general relativity (GR). In GR the causal structure, whilst not given in advance, is part of the solution. After solving Einstein's field equations we know the metric and therefore the causal structure.

Indefinite causal structure would mark a radical departure from previous physics. Many of our basic concepts and modes of thought rely on having definite causal structure. For example, we often think of quantities being conserved (in time) or increasing (in time), and we think of information flowing (in time). We think of entanglement (across space), and, most crucially, we often think of a state (across space) evolving (in time). However, if we have indefinite causal structure there would, in general, be no matter of fact as to whether a particular interval was space-like or time-like and so all of these concepts and modes of thought would be placed under some tension. Nevertheless, one can make a very strong case that quantum gravity (QG) will have indefinite causal structure and so we need to think sufficiently radically to be able to be in a position to deal with this. Most approaches to QG do imagine some form of indefinite causal structure. However, there has been comparatively little thought as to how to really deal with this properly. Generally the conceptual and mathematical tools handed down to us from previous physics are encumbered with ingrained notions of definite causal structure. For example, one might argue that we can model indefinite causal structure by taking a sum over histories, each having its own definite causal structure, but why require each history to have definite causal structure rather than giving up this notion altogether at the fundamental level? We need to be prepared to think radically about this issue. The causaloid formalism offers a way forward here.

A common attitude is that the equations of physics must tell us how to calculate the evolution of physical systems *in time*. If there is indefinite causal structure then

we cannot think in this way. Instead, we adopt the assertion that *a physical theory must correlate recorded data*. This does not commit us to a picture of anything evolving in time. Thus, we might ask what

$$\text{prob}\,(\text{data}_2|\text{data}_1) \tag{3.1}$$

is equal to. If we can deal with all such probabilities for any data then we can say that we have formulated a physical theory (at least that aspect of the theory which can be empirically verified). By thinking about how data might be correlated, we are adopting an operational methodology here. However, this is just a methodology aimed at helping along theory construction. In adopting this approach we do not commit ourselves to operationalism as a fundamental philosophical outlook on the world.

We will now discuss two issues that arise when we think in this way (particularly when there is indefinite causal structure).

3.2.1 Issue 1: the need for a two-step approach

The first issue is the question of when we have sufficient information to be able to make a prediction. Though this is often not appreciated, physical theories attempt to answer only a very small fraction of the possible questions about the world one might put to them. To see this, consider a spin-$\frac{1}{2}$ particle subjected to three sequential spin measurements. The probability that spin up is seen at the second position given that spin up was seen at the first position, and given that the angles chosen were θ_1 and θ_2 (in the first and second positions), can be written

$$\text{prob}(+_2|+_1, \theta_1, \theta_2). \tag{3.2}$$

This probability can be calculated using quantum theory (QT) (and is equal to $\cos^2\left[(\theta_2 - \theta_1)/2\right]$). We can say that this probability is well defined. This is an example of a question that the theory does answer. But now consider the probability that spin up is seen at the third position given that spin up was seen at the first position, and given that the angles chosen were θ_1 and θ_3 (in the first and third positions). We can write this probability as

$$\text{prob}(+_3|+_1, \theta_1, \theta_3). \tag{3.3}$$

Note that we are not given any information about the second spin measurement. This is not part of the conditioning. Under these circumstances we cannot use quantum theory to calculate this probability. This probability is not well defined. This is a question that QT does not answer, and neither should it. Indeed, generically QT does not answer most questions. This is true of physical theories in general (even deterministic ones). In GR, for example, we can make predictions only about data

that may be recorded in some region R_2 given data in region R_1 if R_2 is a domain of dependence of R_1.

The key difference between the two situations in the spin example is to do with the causal structure. In the first case one measurement immediately precedes the other, whereas in the second case there is a gap in time for which we have no information. In order to know whether the probability is well defined or not we need to know what causal situation pertains. If we have definite causal structure then we can refer to it and know whether we are in the rather special type of situation in which we can actually make a prediction. However, if we have indefinite causal structure then we do not know how to proceed.

No doubt there will still be certain conditional probabilities that are well defined even if we go beyond quantum theory and have indefinite causal structure. One way we might deal with this is to mathematize the question. Thus we want a formalism involving two steps:

Step 1 We have a mathematical condition that is satisfied if and only if a probability is well defined.

Step 2 In the case in which the condition in step 1 is satisfied, we have a formula for calculating the probability.

The standard picture with definite causal structure is actually an example of this form. Thus, we have the theory of domains of dependence that tell us whether we can make predictions about some region R_2 on the basis of data in region R_1 by looking at the causal structure. However, we can imagine more general ways in which we might implement this two-step approach that do not explicitly refer to causal structure (at least as the latter is usually conceived).

3.2.2 Issue 2: the need for F-locality

The second issue is very much related to the first. Imagine we want to calculate probabilities pertaining to some arbitrary space-time region R. This space-time region may be of any shape and may be disconnected (insomuch as we have a notion of connection in the absence of definite causal structure). For example we may want to know what the probability of seeing a certain outcome in R is, given that we performed certain measurements in R and saw certain other outcomes in R. In the standard formulation of QT we have a state across space evolving in time. Imagine that R consists of two disconnected parts that are time-like separated. To make a prediction (to say whether the probability is well defined and, if so, what it is equal to) we need to evolve the quantum state through intermediate times. Therefore we necessarily need to refer to mathematical objects and (implicitly) data that do not pertain to R in the evolving state picture. As we will see, this is also the case

in other types of formulation of physical theories (such as histories approaches). If we have some well-defined causal structure then we can use that to tell us what other region, besides R, we need to be considering in order to implement the mathematical machinery of the physical theory. However, this option is not open to us if we have indefinite causal structure. In that case the only clean approach is to insist that, in making predictions for R, we refer only to mathematical objects pertaining to R (for, if not, what do we consider?). This seems like a useful idea and deserves a name – we will call it formalism locality (or F-locality).

A formulation of a physical theory is F-local if, in using it to make statements (using the two-step approach) about an arbitrary space-time region R, we need only refer to mathematical objects pertaining to R.

It is possible that a given physical theory can be formulated in different ways. F-locality is a property of the way the physical theory is formulated rather than of the theory itself. It is possible that any theory admits a formulation that is F-local. In the case in which there is a definite causal structure we may be able to provide both F-local and not-F-local formulations of a theory. However, if there is indefinite causal structure then it seems likely that any fundamental formulation of the theory will necessarily be F-local.

There are two motivations for attempting to formulate theories in an F-local fashion:

1. We do not need to refer to some definite causal structure to decide what other region (besides the region under consideration) to consider.
2. It is more efficient to consider only mathematical objects pertaining to the given region.

Both these reasons are worth bearing in mind when evaluating formulations that are not F-local.

3.3 How standard formulations of physical theories are not F-local

There are, perhaps, three ways in which physical theories have been formulated to date.

1. The state-evolving-in-time picture.
2. Histories formulations.
3. The local-equations approach.

None of these is F-local (the third case is a little more debatable), as we will now see.

In the state-evolving-in-time picture the state is specified at some initial time and it then evolves according to some equations. Imagine we want to make a statement about a space-time region, R, consisting of two disconnected parts that are

separated in both space and time. To do this we take a state defined across enough of space to encompass both spatial regions and evolve it through enough time to encompass the two regions. Hence, we need to refer to mathematical objects pertaining to a region of space-time R' that includes both spatial and temporal regions that are not part of R.

In histories formulations we consider the entire history from some initial to some final time. The physical theory makes statements about such entire histories (the path-integral formulation of quantum theory is one example). If we are interested only in some particular region R then, clearly, in a history formulation, we need to make reference to mathematical objects that do not pertain to R and so the formulation is not F-local. One might claim that, since we can take the history across all of space-time, we do not need to refer to any definite causal structure to decide what region to consider – we simply take everything. Even if this does work, it is still more efficient to aim at an F-local formulation. In practice, we always take our histories over some limited time interval. Indeed, in the absence of a solution, we may not know the nature of "all of space-time"and so it is difficult to know how to give a histories formulation of the theory.

An example of the local-equations approach is Maxwell's equations. Such equations constitute a set of local statements about the infinitesimal regions making up our region R. To actually make a prediction for region R we need to use these local statements appropriately. Typically this involves solving the equations with boundary conditions on a boundary that is in the causal past of all of R. Hence, we need to consider a region bigger than R. There may be other ways to utilize local equations to make predictions about arbitrary regions that do not require consideration of a larger region. It is clear, in any case, that a local-equations formulation is not explicitly F-local as defined above because it does not come equipped with an F-local technique for making predictions for arbitrary regions. There is one sense in which local equations clearly go against the spirit of F-locality. A local equation relates quantities in regions that are infinitesimally displaced from one another. The property of being infinitesimally displaced relates to causal structure. If the causal structure is indefinite then it is not clear that we can retain this notion (this is one reason that we may expect whatever plays the role of space-time in a theory of QG to be discrete rather than continuous).

3.4 An outline of the causaloid framework

It is not clear that physical theories can be formulated in an F-local fashion. In [7, 8] a framework for probabilistic theories with indefinite causal structure was given. This framework provides a way of explicitly formulating theories in an F-local

fashion. Here we will give a bare-bones outline of this framework. In the next section we will indicate how the QT of interacting qubits can be formulated in the framework.

3.4.1 Data and regions

Imagine that all the data collected during an experiment are recorded on cards as triples (x, F_x, Y_x). Here x is some recorded data taken as representing space-time location, F_x is some choice of experiment (knob setting) at x, and Y_x is the outcome of some experiment at x. For example, x might be recorded from a GPS system, F_x could be the angle at which a Stern–Gerlach apparatus is set, and Y_x could be the outcome of the spin measurement. During a typical experiment, data will be recorded at many space-time locations. At the end of one run of the experiment we will collect a stack of cards. Since we are interested in probabilities, we can imagine running the experiment many times so we can obtain relative frequencies.

Since x constitutes recorded data, it will be discrete. Therefore, we can suppose that space-time is discrete and comprised of elementary regions R_x. We do not assume any a-priori causal structure on the x. An arbitrary region R_1 consists of some set of elementary regions R_x,

$$R_1 = \bigcup_{x \in \mathcal{O}_1} R_x. \tag{3.4}$$

We let F_1 denote the list of knob settings F_x for $x \in \mathcal{O}_1$ and Y_1 denote the list of outcomes for $x \in \mathcal{O}_1$. F_1 denotes the choices made in R_1 and Y_1 denotes the outcomes in R_1 (sometimes we will use the longhand notation F_{R_1} and Y_{R_1} for F_1 and Y_1, respectively).

We can ask what

$$\text{prob}(Y_2|Y_1, F_1, F_2) \tag{3.5}$$

is equal to. That is, what is the probability of seeing outcomes Y_2 in region R_2 given that we chose F_1 in region R_1, chose F_2 in R_2, and saw outcomes Y_1 in region R_1 (this is a more sophisticated version of (3.1))? Here we are trying to make statements about region $R_1 \cup R_2$. We will now outline how we go about doing this in an F-local way for $R_1 \cup R_2$.

3.4.2 p-Type vectors and r-type vectors

Let V be the union of all elementary regions. We will assume that the probabilities

$$\text{prob}(Y_V|F_V) \tag{3.6}$$

are well defined (we are glossing over subtle points that are covered in [7]). We can write

$$\text{prob}(Y_V|F_V) = \text{prob}(Y_{R_1}, Y_{V-R_1}|F_{R_1}, F_{V-R_1}). \tag{3.7}$$

We will now label each possible (Y_{R_1}, F_{R_1}) combination in region R_1 with $\alpha_1 \in \Upsilon_1$. This label runs over all possible (outcome, choice) combinations in region R_1. We write

$$p_{\alpha_1} = \text{prob}(Y_{R_1}^{\alpha_1}, Y_{V-R_1}|F_{R_1}^{\alpha_1}, F_{V-R_1}). \tag{3.8}$$

We can regard (Y_{V-R_1}, F_{V-R_1}) in $V - R_1$ as a kind of generalized preparation for region R_1 (it is generalized since it pertains to both the future and the past insofar as those concepts have meaning). Associated with each generalized preparation is a state. We define the state for region R_1 to be that thing which is represented by any mathematical object that can be used to calculate an arbitrary probability p_{α_1}. Clearly the object

$$\begin{pmatrix} \vdots \\ p_{\alpha_1} \\ \vdots \end{pmatrix}, \quad \alpha_1 \in \Upsilon_1, \tag{3.9}$$

suffices (since it simply lists all the probabilities). However, in general, we expect that this is much more information than necessary. In general, in physical theories, all quantities can be calculated from a subset of quantities. We call this *physical compression*. In our particular case we expect that a general probability p_{α_1} can be calculated from a subset of these probabilities. We will restrict ourselves to linear physical compression (where the probabilities are related by linear equations). We set

$$\mathbf{p} = \begin{pmatrix} \vdots \\ p_{k_1} \\ \vdots \end{pmatrix}, \quad k_1 \in \Omega_1 \subseteq \Upsilon_1, \tag{3.10}$$

such that a general probability p_{α_1} can be calculated from the p_{k_1} (with $k_1 \in \Omega_1$) by a linear equation

$$p_{\alpha_1} = \mathbf{r}_{\alpha_1} \cdot \mathbf{p}. \tag{3.11}$$

We chose the fiducial set Ω_1 of labels such that there is no other choice with smaller $|\Omega_1|$ (this means that every probability in $\mathbf{p}$ is necessary in the specification of the state). There may be many possible choices for the fiducial set Ω_1. We simply choose one (for each region) and stick with it. We do not lose generality by imposing linearity here. In the worst case $\Omega_1 = \Upsilon_1$. It is possible that non-linear compression is more efficient. However, for probabilities this is not the case as long

as one can form arbitrary mixtures. In particular, in QT (and classical probability theory) linear compression is optimal (so long as we allow mixed states rather than restrict ourselves to pure states).

Associated with each region R_1 is a real vector space of dimension $|\Omega_1|$. Further, associated with each (Y_1, F_1) combination there is an **r**-type vector (which lives in a dual space to the **p**-type vectors representing the states), which we can write as

$$\mathbf{r}_{(Y_1,F_1)}(R_1) \tag{3.12}$$

or $\mathbf{r}_{\alpha_1}$ for short. It is these **r**-type vectors that we use in real calculations. The state vector, **p**, is akin to scaffolding – it can be dispensed with once the structure of the **r**-type vectors is in place, as we will see shortly.

3.4.3 The causaloid product

If we have two disjoint regions R_1 and R_2, then we can consider the region $R_{12} \equiv R_1 \cup R_2$ as a region in its own right. We can denote the outcome and knob settings for R_{12} as $Y_1 \cup Y_2$ and $F_1 \cup F_2$ (perhaps we are slightly abusing the $\cup$ notation here). Since $R_1 \cup R_2$ is a region in its own right, we will have **r**-type vectors associated with each (outcome, choice) combination in this region also.

We can label the (outcome, choice) pairs $(Y_1 \cup Y_2, F_1 \cup F_2)$ with $\alpha_1\alpha_2 \in \Upsilon_1 \times \Upsilon_2$ (where $\times$ denotes the Cartesian product of ordered pairs taken from the two sets). We also have the fiducial set Ω_{12} for this region. There is an important theorem – namely that it is always possible to choose $\Omega_{12} \subseteq \Omega_1 \times \Omega_2$. We will use the notation $l_1l_2 \in \Omega_1 \times \Omega_2$ and $k_1k_2 \in \Omega_{12}$.

In the case that $\Omega_{12} = \Omega_1 \times \Omega_2$ we have no extra physical compression when two regions are considered together. However, if $\Omega_{12} \subset \Omega_1 \times \Omega_2$ then there is an extra physical compression (second-level compression) for the composite region over and above the physical compression (first-level compression) for each component region considered separately. Physically, this non-trivial case corresponds to *causal adjacency* such as when a qubit passes through two sequential regions with no gap in between.

Second-level compression is quantified by a matrix $\Lambda^{k_1k_2}_{l_1l_2}$ (which depends on the composite region under consideration) such that

$$\mathbf{r}_{\alpha_1\alpha_2}\big|_{k_1k_2} = \sum_{l_1l_2 \in \Omega_1 \times \Omega_2} \mathbf{r}_{\alpha_1}\big|_{l_1} \mathbf{r}_{\alpha_2}\big|_{l_2} \Lambda^{k_1k_2}_{l_1l_2}, \tag{3.13}$$

where $\mathbf{r}_{\alpha_1}\big|_{l_1}$ denotes the l_1 component of $\mathbf{r}_{\alpha_1}$. We write

$$\mathbf{r}_{\alpha_1\alpha_2} = \mathbf{r}_{\alpha_1} \otimes^{\Lambda} \mathbf{r}_{\alpha_2}, \tag{3.14}$$

where the components are given in (3.13). Accordingly, we have defined a new type of product denoted by $\otimes^{\Lambda}$. We call this the *causaloid product*. It is the sought-after unification of the various products $A \otimes B$, AB, and $[A?B]$ from QT mentioned in the introduction (though in a more general framework). The form of the product is the same regardless of the causal structure. However, since the Λ matrix can differ for different composite regions, we can still encode the different products in this one product.

If we have a region regarded as being composed of more than two regions then we can generalize the above ideas appropriately (calling on, in general, a matrix of the form $\Lambda^{k_1 k_2 k_3 \ldots}_{l_1 l_2 l_3 \ldots}$).

3.4.4 The two-step approach in the causaloid framework

We note that, using Bayes' rule,

$$\text{Prob}(Y_2|Y_1, F_1, F_2) = \frac{\text{Prob}(Y_2, Y_1|F_1, F_2)}{\sum_{X_2 \sim F_2} \text{Prob}(X_2, Y_1|F_1, F_2)} \tag{3.15}$$

$$= \frac{\mathbf{r}_{(Y_2 \cup Y_1, F_1 \cup F_2)} \cdot \mathbf{p}}{\sum_{X_2 \sim F_2} \mathbf{r}_{(X_2 \cup Y_1, F_1 \cup F_2)} \cdot \mathbf{p}}, \tag{3.16}$$

where the notation $X_2 \sim F_2$ denotes all outcomes in R_2 that are consistent with the choice F_2 in R_2. In order that the probability $\text{Prob}(Y_2|Y_1, F_1, F_2)$ be well defined, it must depend only on the given conditioning in $R_1 \cup R_2$ and not depend on what happens outside this region. The state, on the other hand, is associated with some generalized preparation in $V - R_1 - R_2$. Hence, this probability is well defined only if there is no dependence on the state $\mathbf{p}$ (strictly we should have included the conditioning in $V - R_1 - R_2$ and then shown that it is irrelevant if there is no dependence on $\mathbf{p}$). We can use this observation to implement a two-step approach.

1. $\text{Prob}(Y_2|Y_1, F_1, F_2)$ is well defined if and only if

$$\mathbf{r}_{(Y_2 \cup Y_1, F_1 \cup F_2)} \text{ is parallel to } \sum_{X_2 \sim F_2} \mathbf{r}_{(X_2 \cup Y_1, F_1 \cup F_2)} \tag{3.17}$$

   (since this is the necessary and sufficient condition for there being no dependence on $\mathbf{p}$).
2. If these vectors are parallel, then the probability is given by

$$\text{Prob}(Y_2|Y_1, F_1, F_2) = \frac{\left|\mathbf{r}_{(Y_2 \cup Y_1, F_1 \cup F_2)}\right|}{\left|\sum_{X_2 \sim F_2} \mathbf{r}_{(X_2 \cup Y_1, F_1 \cup F_2)}\right|}. \tag{3.18}$$

We see that this two-step approach does not require us to refer to some given definite causal structure (at least as the latter is usually conceived).

We see that the formulation is F-local since, to make predictions for an arbitrary space-time region (in this case the region $R_1 \cup R_2$), we need only consider mathematical objects pertaining to this region (the **r** vectors). Strictly speaking, we need to be sure we can calculate the **r** vectors for arbitrary regions without referring to mathematical objects pertaining to other regions to assert that the formulation is fully F-local. This will be addressed below.

3.4.5 The causaloid

We can calculate **r** vectors for an arbitrary region by starting from the **r** vectors for the elementary regions comprising that region and using the causaloid product. Hence, we can calculate any **r**-vector if we know

(1) all the vectors $\mathbf{r}_{\alpha_x}$ (which can be regarded as a matrix $\Lambda_{\alpha_x}^{k_x}$) for all elementary regions, R_x, and
(2) all the matrices $\Lambda_{l_x l_{x'} l_{x''} \ldots}^{k_x k_{x'} k_{x''} \ldots}$ with $x, x', x'', \ldots \in \mathcal{O}_1$, for all $\mathcal{O}_1$ with $|\mathcal{O}_1| \geq 2$ (since these pertain to composite regions).

This constitutes a tremendous amount of information (the number of matrices is exponential in the number of elementary regions and the size of these matrices grows with the size of the region they pertain to). However, we can apply physical compression by finding relationships between these matrices (we call this *third-level compression*). After applying physical compression, we have a smaller set of matrices from which all the others can be calculated. We call this smaller set, augmented by a set of rules for implementing decompression, *the causaloid* and denote it by Λ. Since we want the framework to be F-local, we require that, in applying decompression to obtain the matrix for a region R_1, we need use only matrices pertaining to regions $\tilde{R}_1 \subseteq R_1$.

While we have not shown how to calculate the causaloid in general, it has been shown how to do so for the classical probabilistic theory of pairwise-interacting classical bits and the QT of pairwise-interacting qubits. We will outline, in the next section, how this works in the quantum case. The classical case is very similar (though we will not outline it here).

3.5 Formulating quantum theory in the causaloid framework

In this section we will content ourselves with simply describing how the QT of pairwise-interacting qubits can be formulated in the causaloid framework without deriving any of the results. Universal quantum computation can be implemented with pairwise-interacting qubits and so arbitrary quantum systems can be simulated to arbitrary accuracy (similar comments apply in the classical case). Hence the

case we are studying is more than just an example. It demonstrates (with some appropriate qualifications) that QT in general can be formulated in this framework.

Assume we have a large number of qubits moving to the right labeled (from left to right) $u = 1, 2, 3, \ldots$ and a large number of qubits moving to the left labeled (from right to left) $v = 1, 2, 3, \ldots$. If this is plotted against time then we will have a diamond-shaped lattice with each vertex corresponding to the interaction of a right-moving qubit with a left-moving qubit. We can label these vertices with $x \equiv uv$ (the Cartesian product of u and v is denoted by uv). They correspond to our elementary regions R_{uv}.

We imagine that, at each vertex, the two qubits pass through a box, which implements a general measurement. The box has a knob that is used to set F_{uv} and a display panel recording the outcome, Y_{uv}. As before, we can label all such pairs with α_{uv}. For a general quantum measurement, an (outcome, choice) pair is associated with a superoperator \$ (superoperators are trace non-increasing maps on density operators that take allowed states to allowed states). In this case, we have a superoperator $\$_{\alpha_{uv}}$ associated with α_{uv}. In QT we can write a general superoperator on two qubits such as this as a sum of the tensor product of a fiducial set of superoperators acting on each qubit separately:

$$\$_{\alpha_{uv}} = \sum_{k_u k_v \in \Omega^2 \times \Omega^2} \Lambda_{\alpha_{uv}}^{k_u k_v} \$_{k_u} \otimes \$_{k_v}. \tag{3.19}$$

Remarkably, we can do this only if we have complex (rather than real or quaternionic) Hilbert spaces supporting the superoperators. The set $\$_{k_u}$ for $k_u \in \Omega^2$ is a fiducial spanning set of superoperators for the qubit (the superscript denotes that this is a qubit having Hilbert-space dimension 2). We have $|\Omega^N| = N^4$, so for a qubit we have $|\Omega|^2 = 2^4$. This is the number of linearly independent superoperators needed to span the space of superoperators for a qubit. For reasons that will be clear later, we choose $\$_1 = I$, where 1 is the first element of Ω^2 and I is the identity superoperator. We can solve (3.19) to find the Λ matrix for each elementary region R_{uv}.

Now consider a single right-moving qubit as it goes from vertex (u, v) to the next vertex $(u, v + 1)$. Assume that it is subject to $\$_{l_u} \otimes \$_{l_v}$ at the first vertex and $\$_{l'_u} \otimes \$_{l'_{v+1}}$ at the next, with $l_u, l'_u, l_v, l'_{v+1} \in \Omega^2$. Insofar as the right-moving qubit, u, is concerned, we can ignore the left-moving qubits with which it interacts (since the two superoperators just given factorize). The effective superoperator acting on qubit u is $\$_{l'_u} \circ \$_{l_u}$. But this is a superoperator belonging to the space of superoperators acting on a single qubit and can, hence, be expanded in terms of the linearly independent fiducial set

$$\$_{l'_u} \circ \$_{l_u} = \sum_{k'_u k_u \in \{1\}\times\Omega^2} \Lambda^{k'_u k_u}_{l'_u l_u} \$_{k'_u} \circ \$_{k_u} \tag{3.20}$$

since we have selected $\$_1$ to be the identity. We can solve this this equation for the matrix $\Lambda^{k'_u k_u}_{l'_u l_u}$.

This generalizes to more than two sequential vertices in the obvious way. For three sequential vertices we have

$$\$_{l''_u} \circ \$_{l'_u} \circ \$_{l_u} = \sum_{k''_u k'_u k_u \in \{1\}\times\{1\}\times\Omega^2} \Lambda^{k''_u k'_u k_u}_{l''_u l'_u l_u} \$_{k''_u} \$_{k'_u} \circ \$_{k_u} \tag{3.21}$$

and so on.

It can be shown that, for three sequential vertices,

$$\Lambda^{k''_u k'_u k_u}_{l''_u l'_u l_u} = \sum_{n'\in\Omega^2} \Lambda^{k''_u n'_u}_{l''_u l'_u} \Lambda^{k'_u k_u}_{n'_u l_u}. \tag{3.22}$$

For four sequential vertices,

$$\Lambda^{k'''_u k''_u k'_u k_u}_{l'''_u l''_u l'_u l_u} = \sum_{n''\in\Omega^2,\ n'\in\Omega^2} \Lambda^{k'''_u n''_u}_{l'''_u l''_u} \Lambda^{k''_u n'_u}_{n''_u l'_u} \Lambda^{k'_u k_u}_{n'_u l_u} \tag{3.23}$$

and so on. The derivation of these equations relies only on the combinatorics of how the labels combine rather than on any particular details of QT. The same equations are found in the treatment of interacting classical bits.

We may have need of the matrix $\Lambda^{k_u}_{l_u}$ (where $l_u, k_u \in \Omega^2$) for a single vertex for a right-moving qubit. Since

$$\$_{l_u} = \sum_{k_u\in\Omega^2} \Lambda^{k_u}_{l_u} \$_{k_u} \tag{3.24}$$

we have

$$\Lambda^{k_u}_{l_u} = \delta^{k_u}_{l_u}. \tag{3.25}$$

For left-moving qubits we have (3.20)–(3.25) but with v replacing u.

As will be described, the composite-region Λ matrices for all situations that do not involve sequential vertices on the same qubit are given by multiplying the Λ matrix components for different clumps of vertices. This means that the causaloid is given by

$$\Lambda = (\Lambda^{k_u k_v}_{\alpha_{uv}} \ \forall\, uv,\ \Lambda^{k'_u k_u}_{l'_u l_u} \ \forall\, \text{RSV},\ \Lambda^{k'_v k_v}_{l'_v l_v} \ \forall\, \text{LSV};\ \text{clumping method}), \tag{3.26}$$

where (LSV) RSV stands for pairs of sequential vertices on (left-) right-moving qubits. The clumping method allows us to calculate the Λ matrix for an arbitrary region R_1 with $x\ (= uv) \in \mathcal{O}_1$ from this causaloid as follows.

1. For each qubit (both left- and right-moving) circle all complete groups of sequential vertices (call these *clumps*) in $\mathcal{O}_1$. There must be a gap of at least one vertex between each clump and the next for any given qubit.
2. Calculate the Λ matrix components for each circled clump for each qubit using (3.22) or one of its generalizations (for a clump of one vertex use (3.25) and for a clump of two vertices take $\Lambda^{k'_u k_u}_{l'_u l_u}$ directly from the specification of the causaloid (3.26)).
3. Multiply together all Λ matrix components for all circled clumps (note that we are multiplying components rather than performing matrix multiplication). This gives the components of the Λ matrix for R_1.

We note that the clumping method respects F-locality since, in calculating the Λ matrix for R_1, we use only Λ matrices pertaining to regions $\tilde{R}_1 \subseteq R_1$.

It is worth examining the causaloid given in (3.26) a little more. First we note that we are required to specify only a tiny subset of the exponential number of possible Λ matrices – there is a tremendous amount of third-level compression. Second, we note the symmetry property that, according to (3.19) and (3.20), the Λ matrices of each type are the same. Hence, we can actually specify the causaloid by

$$\Lambda = (\Lambda^{k_1 k_1}_{\alpha_{11}}, \Lambda^{k'_1 k_1}_{l'_1 l_1}; \text{symmetry, clumping method}), \tag{3.27}$$

where *symmetry* denotes the property just noted and $\Lambda^{k'_1 k_1}_{l'_1 l_1}$ is one instance of the Λ matrix for a pair of right- (or left-) sequential vertices.

Given this causaloid, we can employ the standard techniques of the causaloid framework (the causaloid product and the two-step approach) to calculate whether an arbitrary probability is well defined and, if so, what it is equal to. In this sense we can say that this causaloid fully specifies the QT of interacting qubits. In particular, note that we separate out the specification of the theory (the causaloid will be different for different physical theories) from the way the causaloid is used to make predictions (it is used in the same way for any physical theory).

The causaloid formulation of QT treats arbitrary regions on an equal footing. In this it is similar to the time-symmetric approach of Aharanov and co-workers (particularly the latest version due to Aharanov, Popescu, Tollaksen, and Vaidman in [9], which allows states and measurements to be defined for arbitary regions) and the general boundary formulation of quantum theory due to Oeckl [10]. Of related interest are the quantum causal histories approach of Markopoulou [11] and the quantum causal networks of Leifer [12].

3.6 The road to quantum gravity

In order to formulate a theory of QG we need to have a framework that is hospitable to such a theory in the first place. We expect that QG will be a probabilistic theory

with indefinite causal structure. The causaloid framework admits such theories (this does not imply that QG certainly fits within the framework – but at least it is not ruled out from the outset). The most satisfying way to obtain QG in this (or any) framework would be to derive it from a set of well-motivated principles (such as a suitably generalized equivalence principle). It may be that appropriate principles will carry over from QT and GR (indeed, the equivalence principle is true in Newtonian gravity). Therefore, a careful study of QT and GR may be the best way of coming up with such principles.

One particular route that may be taken to finding QG is illustrated in the following diagram:

$$\begin{array}{ccc} \text{QT} & \longrightarrow & \text{QG} \\ \uparrow & & \uparrow \\ \text{CProbT} & \longrightarrow & \text{ProbGR} \end{array}$$

where CProbT is classical probability theory (of interacting classical bits, for example), QT is quantum theory (of interacting qubits, for example), and ProbGR is an appropriately formulated version of general relativity in the case in which we have arbitary probabilistic ignorance of the values of certain measurable quantities. We will elaborate on this below. The vertical arrows represent a kind of *quantization*. The horizontal arrows represent what we might call *GR-ization*. In quantizing from CProbT to QT we need only alter the structure of the $\Lambda_{\alpha_{uv}}^{k_u k_v}$ matrices for the elementary regions (this is what might be regarded as the local structure) by replacing that structure which corresponds to a classical probability simplex with a structure that corresponds to the Bloch sphere of the qubit. The structure above this, for composite regions, is basically constructed in the same way in CProbT and QT. ProbGR has not yet been satisfactorily formulated. However, we can expect that, in the GR-ization process from CProbT to ProbGR, the local structure associated with the classical probability simplices will survive for the elementary regions but the structure associated with composite regions will be different (since the causal structure is not fixed). This suggests that we may be able to get a theory of QG by applying quantization essentially at the local level of the elementary regions and GR-ization at the level of the composite regions. If quantization and GR-ization do not interfere with each other too much then the diagram above might not be too misleading.

This approach depends on having a suitable formulation of ProbGR. An obvious way to give a probabilistic formulation of GR is to have some probabilistic distribution over the 3-metric specified on some initial space-like hypersurface and then evolve the distribution employing a canonical formulation of GR. This is unsatisfactory since (a) it is not F-local, (b) we cannot deal with arbitrary probabilistic ignorance about measurable quantities, and (c) the time label for the space-like

hypersurface is not an observable and so it is not clear that the numbers we are calling "probabilities" represent ignorance about something measurable. Another approach is to consider a probabilistic distribution over solutions for the metric for all of space-time. This is problematic since (a) it is manifestly not F-local and (b) it is not clear how the differently weighted solutions match up from the internal point of view of somebody who may be making measurements and so, once again, it is unclear whether the "probabilities" represent ignorance about something measurable. Rather than taking either of these approaches, it seems that we need to build up ProbGR from scratch using F-locality and, maybe, the causaloid formalism as guidance. Such a theory contains no new empirical content over standard GR. However, it is possible that the natural mathematical formulation of ProbGR will look very different from standard GR.

3.7 Conclusions

Many standard notions in QT require reference to some definite causal structure. For example the notion of entanglement requires two space-like separated systems, and the notion of information flow requires a sequence of immediately sequential time-like regions. When we embed QT into the causaloid framework these notions become special cases of a much richer structure. Entanglement is supported by the tensor product of QT, but, in the causaloid framework, we have the causaloid product which allows us to talk about joint properties of any two regions regardless of their causal relationship. Information flow is supported by the standard product $\hat{A}\hat{B}$ between sequential time-like separated regions. In the causaloid framework we have, again, the causaloid product. In quantum circuit diagrams we draw wires between boxes denoting the path of the qubit. Any two boxes either do, or do not, have a wire between them. In the causaloid framework we have a Λ matrix (by which the causaloid product is defined). For every pair of boxes (or elementary regions) there is a Λ matrix between the two boxes. This richer structure is likely to help in developing a theory of QG since it provides a way round requiring that the causal structure be definite.

The principle that it should be possible to give an F-local formulation may prove to be powerful in theory construction. It is encouraging that QT can be formulated in an F-local fashion. Not only does this add to the list of different ways in which QT can be formulated but also it provides encouragement that a theory of QG may share structural similarities with QT.

The next step on the road to QG, if this approach is pursued, is to construct ProbGR. In tackling the problem of constructing ProbGR we are likely to encounter many of the same difficulties as would be encountered in constructing a fully

fledged theory of QG. However, we know that ProbGR is empirically equivalent to GR (just with arbitrary probabilistic ignorance added) and so we fully expect that this theory exists.

References

[1] J. Polchinski, *String Theory*, vols. 1 and 2 (Cambridge: Cambridge University Press, 1998).
[2] C. Rovelli, *Quantum Gravity* (Cambridge: Cambridge University Press, 2004); T. Thiemann, *Lectures on Loop Quantum Gravity* (Berlin: Springer, 2003); L. Smolin, An invitation to loop quantum gravity, hep-th/0408048 (2004); L. Smolin, Introduction to quantum gravity, PIRSA:C06001 (2006).
[3] R. Penrose, On gravity's role in quantum state reduction, *Gen. Rel. Grav*. **28**, 581 (1996).
[4] R. D. Sorkin, Causal sets: discrete gravity, qr-qc/0309009 (2003); R. D. Sorkin, Is a past-finite causal order the inner basis of spacetime?, PIRSA:05090001 (2005).
[5] A. Doering and C. Isham, What is a thing?: *topos* theory in the foundations of physics, arXiv:0803.0417 (2008).
[6] J. B. Hartle, Generalizing quantum mechanics for quantum gravity, gr-qc/0510126 (2005).
[7] L. Hardy, Probabilistic theories with dynamic causal structure: a new framework for quantum gravity, gr-qc/0509120 (2005).
[8] L. Hardy, Towards quantum gravity: a framework for probabilistic theories with non-fixed causal structure, *J. Phys. A* **40**, 3081 (2007); L. Hardy, Operationalism and quantum gravity, PIRSA:07060042 (2007).
[9] Y. Aharanov, S. Popescu, J. Tollaksen, and L. Vaidman, Multiple-time states and multiple-time measurements in quantum mechanics, arXiv:0712.0320 (2007).
[10] R. Oeckl, General boundary quantum field theory: foundations and probability interpretation, hep-th/0509122 (2005); R. Oeckl, The general boundary formulation of quantum mechanics: motivations and elementary examples, PIRSA:04110000 (2004).
[11] F. Markopoulou, Quantum causal histories, hep-th/9904009 (1999).
[12] M. Leifer, Quantum causal networks, PIRSA:06060063 (2006).

Part II

Quantum probability

4

Bell's inequality from the contextual probabilistic viewpoint

Andrei Khrennikov

4.1 Introduction

4.1.1 Quantum information and quantum foundations

Quantum information science is about the processing of information by the exploitation of some distinguishing features of quantum systems, such as electrons, photons, ions. In recent years a lot has been promised in the domain of quantum information. In quantum computing it was promised that NP-problems would be solved in polynomial time. In quantum cryptography there were claims that protocols would have practically 100% security. At the moment it is too early to say anything definitive regarding the final results of this great project.

In quantum computing a few quantum algorithms and developed devices, "quantum pre-computers" with a few quantum registers, were created. However, difficulties could no longer be ignored. For some reason it was impossible to create numerous quantum algorithms that could be applied to various problems. Up to now the whole project is based on two or three types of algorithm, and among them one, namely, the algorithms for prime factorization, might be interesting for real-world application. There is a general tendency to consider this situation with quantum algorithms as an occasional difficulty. But, as the years pass, one might start to think that there is something fundamentally wrong. The same feelings are induced by developments in quantum hardware. It seems that the complexity of the problem of creation of a device with a large number N of quantum registers increases extremely non-linearly with increasing N.

In quantum cryptography the situation is opposite to that of quantum computing. There were tremendous successes in the development of technologies for production and transmission of quantum information, especially pairs of entangled

Philosophy of Quantum Information and Entanglement, ed. A. Bokulich and G. Jaeger. Published by Cambridge University Press.

photons.[1] On the other hand, the claim regarding 100% security of quantum protocols is far from being totally justified.

Any careful analysis of this situation implies immediately that the whole project of "*quantum information*" should be based on more solid foundations. We recall that quantum mechanics by itself is a huge building with a sandy foundation – the orthodox Copenhagen interpretation. On the one hand, an advanced mathematical formalism (calculus of probabilities in the complex Hilbert space) giving predictions that are supported by all existing experimental data was created. On the other hand, it is still unclear why this formalism works so well and, moreover, it is not clear what it really predicts, because in the orthodox Copenhagen interpretation (which is the conventional interpretation) quantum mechanics is not about physical reality itself, being rather only about our observations – but of what? All unsolved problems in quantum foundations are essentially amplified in the quantum information project. Problems that were of a purely philosophical interest [1–3] for 100 years have become technological problems of interest to business.

Therefore "*quantum information*" gives a new great opportunity for the reconsideration of quantum foundations, see, e.g., [4]. Whether this chance will be used depends on many scientific, psychological, and market factors. Unfortunately, at the moment there is a tendency to ignore fundamental difficulties and reduce everything to experimental and technological problems. Of course, the development of quantum technologies, in particular, the manipulation of individual quantum systems, is an extremely interesting project. But I hope that it could be done essentially better if quantum computing and cryptography were also considered new tools for testing the foundations of quantum mechanics.

First of all we should come back to the greatest debate of twentieth-century physics, namely the debate between Einstein and Bohr on the *completeness of quantum mechanics* [5]. Nowadays it is commonly accepted that quantum mechanics is complete: the ψ-function provides the most complete description of the state of a quantum system. It is impossible to find a more detailed description of quantum reality – to find a model with hidden variables. This is the basis of the orthodox Copenhagen interpretation and nowadays it is the basis of quantum cryptography. If one were able to find a model with hidden variables that reproduce quantum statistics, then the total security of quantum protocols would be questioned!

In probabilistic terms this is the problem of so-called *irreducible quantum randomness.* In opposition to classical randomness, it is claimed (since von

[1] We emphasize that at the moment photons give the most real basis of quantum cryptography. It is doubtful that there would be created systems for quantum cryptography that would be based on, e.g., electrons. It is not easy to imagine refrigerators with electrons that could be used for transportation of quantum information.

Neumann [1]) that quantum randomness cannot be reduced to conventional ensemble randomness. I think that there should be a very serious discussion on the topic of "quantum foundations in light of quantum information" or "quantum information in light of quantum foundations" [4]. One of the possibilities is to start with Bell's inequality, since its violations play the fundamental role in the foundations of quantum information. It is commonly believed that the experimental violation of Bell's inequality must be interpreted as proof of *action at a distance* or *quantum non-locality.*[2]

4.1.2 Bell's inequality

Bell's inequality has surpassed all records for publications, citations, discussions, and controversies. As an organizer of several international conferences on the foundations of quantum theory, I was really disappointed by stormy debates over Bell's inequality. I could find no problem in quantum foundations which can be compared with Bell's inequality in terms of the intensity of discussions and strength of reactions to opponents' views, [4, 8–19]. I asked myself many times "*What is so special about Bell's inequality?*"

It seems that the main problem stems from the fact that J. S. Bell formulated precisely only his aim: to prove the non-locality of quantum mechanics.[3] However, he did not determine precisely the mathematical probabilistic rules which he used to formalize the problem [6]. This absence of rigor in the mathematical formulation of the problem provides many opportunities for speculation. Each year people present their own (very different) views on the probabilistic structure of the EPR–Bohm experiment and, consequently, totally different physical conclusions.

I have had numerous conversations with outstanding physicists about this situation. It is a rather common opinion that it is pointless to pay attention to probability theory. Typically such a viewpoint is motivated by considering probability as a "physically well-defined quantity." Therefore one need not take care to provide a mathematically rigorous probabilistic formalization. I totally disagree with such a viewpoint. In the same way one might say that physicists need not concern themselves with mathematical models of space and geometries. Fortunately, nobody would say such nonsense nowadays.

[2] In fact, modern formulations of the consequences of the Bell theorem are presented in a trickier way. *Quantum mechanics is incompatible with local realism.* However, is it easy to understand what "non-local realism" could mean at all? In any case non-local realism is not realism in physical space. Therefore I support the initial argument of EPR that precise (anti)correlations give strong support for realism. In such a case the only possible conclusion from the Bell theorem is non-locality. In fact, it was the original viewpoint of J. S. Bell [6, 7].

[3] J. S. Bell was a "non-local realist." The main aim of his investigations may be seen as finding arguments supporting the Bohmian model of quantum mechanics. Thus he wanted to save realism even at the price of locality [6, 7].

Suppose at the moment that we try to work in physics without determining precisely the mathematical models of space and the geometries. It is clear that such an activity would induce permanent debates and even a variety of paradoxes. I think that a similar thing has happened with Bell's inequality, simply with probability, rather than space. Thus the intensity of debates on Bell's inequality is not completely determined by the physical importance of this problem. The fact that the problem was not formulated in a rigorous mathematical framework cannot be neglected. Nevertheless, during the last 40 years physicists have been trying to proceed with Bell's inequalities without describing precisely the probabilistic rules that they used. Moreover, there is a rather common opinion that it is possible to work only with frequencies. Hence, again, the rigorous mathematical description of a corresponding probability model is viewed as unimportant.

In this chapter I shall present a mathematical formalization of the Bell [6] and von Neumann [1] no-go arguments. By proceeding with such a rigorous mathematical framework, the common conclusion regarding the disagreement of the quantum formalism with a classical statistical description is shown to be insufficiently justified.

4.2 Measure-theoretical derivation of Bell-type inequalities

Since our argument includes analysis of standard proofs of the Bell-type inequalities [6, 8], to provide background we first briefly reproduce these well-known proofs.

4.2.1 Bell's inequality

By the Kolmogorov axiomatics [20], see also [21–23] the *probability space* is a triple

$$\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P}),$$

where Ω is an arbitrary set, $\mathcal{F}$ is an arbitrary σ-algebra[4] of subsets of Ω, and $\mathbf{P}$ is a σ-additive measure on $\mathcal{F}$ that yields values in the segment $[0, 1]$ of the real line and normalized by the condition $\mathbf{P}(\Omega) = 1$. *Random variables* on $\mathcal{P}$ are by definition measurable functions

$$\xi : \Omega \to \mathbf{R}. \tag{4.1}$$

Thus $\xi^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-algebra on the real line. Finally, let $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$ be a Kolmogorov probability space. For any pair of random variables $u(\omega)$, $v(\omega)$, their covariation is defined by

[4] In the literature one also uses the terminology σ-field, instead of σ-algebra.

$$\langle u, v\rangle = \mathrm{cov}(u, v) = \int_\Omega u(\omega)v(\omega)\mathrm{d}\mathbf{P}(\omega).$$

We can now reproduce the proof of Bell's inequality in a measure-theoretic framework.

Theorem 4.1. (Bell inequality for covariations) *Let $\xi_a, \xi_b, \xi_c = \pm 1$ be random variables on $\mathcal{P}$. Then Bell's inequality*

$$|\langle \xi_a, \xi_b\rangle - \langle \xi_c, \xi_b\rangle| \leq 1 - \langle \xi_a, \xi_c\rangle \tag{4.2}$$

holds.

Proof. Set $\Delta = \langle \xi_a, \xi_b\rangle - \langle \xi_c, \xi_b\rangle$. By virtue of the linearity of Lebesgue integrals we obtain

$$\Delta = \int_\Omega \xi_a(\omega)\xi_b(\omega)\mathrm{d}\mathbf{P}(\omega) - \int_\Omega \xi_c(\omega)\xi_b(\omega)\mathrm{d}\mathbf{P}(\omega) = \int_\Omega [\xi_a(\omega) - \xi_c(\omega)]\, b(\omega)\mathrm{d}\mathbf{P}(\omega). \tag{4.3}$$

Since

$$\xi_a(\omega)^2 = 1, \tag{4.4}$$

we have

$$|\Delta| = \left|\int_\Omega [1 - \xi_a(\omega)\xi_c(\omega)]\, \xi_a(\omega)\xi_b(\omega)\mathrm{d}\mathbf{P}(\omega)\right| \leq \int_\Omega [1 - \xi_a(\omega)\xi_c(\omega)]\, \mathrm{d}\mathbf{P}(\omega). \tag{4.5}$$

4.2.2 Wigner's inequality

We recall the following simple mathematical result, see Wigner [9].

Theorem 4.2. (Wigner inequality) *Let $\xi_a, \xi_b, \xi_c = \pm 1$ be arbitrary random variables on a Kolmogorov space $\mathcal{P}$. Then the following inequality holds:*

$$\mathbf{P}(\xi_a = +1, \xi_b = +1) + \mathbf{P}(\xi_b = -1, \xi_c = +1) \geq \mathbf{P}(\xi_a = +1, \xi_c = +1). \tag{4.6}$$

Proof. We have

$$\begin{aligned}
&\mathbf{P}(\omega \in \Omega\colon\ \xi_a(\omega) = +1, \xi_b(\omega) = +1)\\
&\quad = \mathbf{P}(\omega \in \Omega\colon\ \xi_a(\omega) = +1, \xi_b(\omega) = +1, c(\omega) = +1)\\
&\qquad + \mathbf{P}(\omega \in \Omega\colon\ \xi_a(\omega) = +1, \xi_b(\omega) = +1, \xi_c(\omega) = -1),
\end{aligned} \tag{4.7}$$

$$\begin{aligned}
&\mathbf{P}(\omega \in \Omega\colon\ \xi_b(\omega) = -1, \xi_c(\omega) = +1)\\
&\quad = \mathbf{P}(\omega \in \Omega\colon\ \xi_a(\omega) = +1, \xi_b(\omega) = -1, c(\omega) = +1)\\
&\qquad + \mathbf{P}(\omega \in \Omega\colon\ \xi_a(\omega) = -1, \xi_b(\omega) = -1, \xi_c(\omega) = +1),
\end{aligned} \tag{4.8}$$

and

$$\begin{aligned}\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=+1,\xi_c(\omega)=+1)\\ =\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=+1,\xi_b(\omega)=+1,\xi_c(\omega)=+1)\\ +\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=+1,\xi_b(\omega)=-1,\xi_c(\omega)=+1).\end{aligned}\tag{4.9}$$

If we add together (4.7) and (4.8) we obtain

$$\begin{aligned}&\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=+1,\xi_b(\omega)=+1)+\mathbf{P}(\omega\in\Omega\colon\ \xi_b(\omega)=-1,\xi_c(\omega)=+1)\\ &\quad=\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=+1,\xi_b(\omega)=+1,\xi_c(\omega)=+1)\\ &\qquad+\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=+1,\xi_b(\omega)=+1,\xi_c(\omega)=-1)\\ &\qquad+\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=+1,\xi_b(\omega)=-1,\xi_c(\omega)=+1)\\ &\qquad+\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=-1,\xi_b(\omega)=-1,\xi_c(\omega)=+1).\end{aligned}\tag{4.10}$$

But the first and third terms on the right-hand side of this equation are just those which when added together make up the term $\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=+1,\xi_c(\omega)=+1)$ (Kolmogorov probability is additive). It therefore follows that

$$\begin{aligned}&\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=+1,\xi_b(\omega)=+1)+\mathbf{P}(\omega\in\Omega\colon\ \xi_b(\omega)=-1,\xi_c(\omega)=+1)\\ &\quad=\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=+1,\xi_c(\omega)=+1)\\ &\qquad+\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=+1,\xi_b(\omega)=+1,\xi_c(\omega)=-1)\\ &\qquad+\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=-1,\xi_b(\omega)=-1,\xi_c(\omega)=+1).\end{aligned}\tag{4.11}$$

By using the non-negativity of probability we obtain the inequality

$$\begin{aligned}&\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=+1,\xi_b(\omega)=+1)+\mathbf{P}(\omega\in\Omega\colon\ \xi_b(\omega)=-1,\xi_c(\omega)=+1)\\ &\quad\geq\mathbf{P}(\omega\in\Omega\colon\ \xi_a(\omega)=+1,\xi_c(\omega)=+1).\end{aligned}\tag{4.12}$$

4.2.3 The Clauser–Horne–Shimony–Holt inequality

Finally, we derive the Clauser–Horne–Shimony–Holt (CHSH) inequality [8].

Theorem 4.3. (CHSH inequality) *Let $\mathcal{P}=(\Omega,\mathcal{F},\mathbf{P})$ be a Kolmogorov probability space and let $\xi_j(\omega)$ and $\xi'_j(\omega)$, $j=1,2$, be random variables such that*

$$|\xi_j(\omega)|\leq 1,\ \ |\xi'_j(\omega)|\leq 1,\ \textit{for almost all } \omega\in\Omega.\tag{4.13}$$

Then the following inequality for correlations holds:

$$\langle\xi_1,\xi'_1\rangle+\langle\xi_1,\xi'_2\rangle+\langle\xi_2,\xi'_1\rangle-\langle\xi_2,\xi'_2\rangle\leq 2.\tag{4.14}$$

Proof. It is easy to prove the following elementary algebraic inequality for numbers bounded by 1:

$$u_1v_1+u_1v_2+u_2v_1-u_2v_2\leq 2.$$

Thus

$$\xi_1(\omega)\xi_1'(\omega) + \xi_1(\omega)\xi_2'(\omega) + \xi_2(\omega)\xi_1'(\omega) - \xi_2(\omega)\xi_2'(\omega) \leq 2. \tag{4.15}$$

Finally, we integrate this inequality with respect to the measure **P**.

4.3 Formalization of rules for correspondence between classical and quantum statistical models

It is assumed that there exists a space of hidden variables Ω representing states of individual physical systems. There is fixed a σ-algebra $\mathcal{F}$ of subsets of Ω. On this space there are defined *classical quantities.* These are measurable functions $\xi\colon \Omega \to \mathbf{R}$ – random variables on the measurable space $(\Omega, \mathcal{F})$.[5] There is also considered to be a space of physical observables O. In the quantum model they are represented by self-adjoint operators, e.g., we can represent O by $\mathcal{L}_s(\mathcal{H})$ – the space of bounded self-adjoint operators in the Hilbert space of quantum states $\mathcal{H}$. We shall distinguish a physical observable and its operator-representative by using symbols a and $\hat{a}$.

The main question concerns the existence of a correspondence between the space of random variables $V(\Omega)$ and the space of quantum observables $\mathcal{L}_s(\mathcal{H})$ – namely the possibility of constructing a map

$$j\colon\ V(\Omega) \to \mathcal{L}_s(\mathcal{H}) \tag{4.16}$$

or a map

$$i\colon\ \mathcal{L}_s(\mathcal{H}) \to V(\Omega) \tag{4.17}$$

that has "natural probabilistic properties" (in general j is not a one-to-one map; its existence does not imply the existence of i and vice versa). The main problem is that physics does not tell us which features such maps should have. There is a large realm for mathematical fantasies, presented in the form of no-go theorems. Let us next recall the history of this problem.

[5] In this framework classical is equivalent to the existence of a *functional representation.* Denote the space of classical quantities by the symbol $V(\Omega)$. This is some space of real-valued (measurable) functions on Ω. The choice of this functional space depends on a model under consideration. For a system whose state is given by the hidden variable ω, the value $\xi(\omega)$ of a classical quantity ξ gives the objective property ξ of this system. We shall not distinguish a classical (physical) quantity and its representation by a random variable.

4.4 Von Neumann postulates on classical–quantum correspondence and the no-go theorem

Any statistical model contains a space of *statistical states*. In a pre-quantum statistical model (which we are looking for) statistical states are represented by probability measures on the space of hidden variables Ω. Denote such a space of probabilities by $S(\Omega)$. This space is chosen depending on a classical statistical model under consideration.[6] In the quantum model statistical states are represented by von Neumann *density operators*. This space is denoted by $\mathcal{D}(\mathcal{H})$. J. von Neumann was the first to present a list of possible features of the classical $\rightarrow$ quantum map j, see [1]:

(VN1) j is a one-to-one map.[7]
(VN2) For any Borel function $f: \mathbf{R} \rightarrow \mathbf{R}$, we have $j(f(\xi)) = f(j(\xi)), \xi \in V(\Omega)$.
(VN3) $j(\xi_1 + \xi_2 + \ldots) = j(\xi_1) + j(\xi_2) + \ldots$ for any any sequence $\xi_k \in V(\Omega)$.[8]

Roughly speaking, von Neumann proved that under conditions (VN1)–(VN3) every operation of statistical averaging on $V(\Omega)$ can be represented as the quantum trace-average on $\mathcal{L}_s(\mathcal{H})$ corresponding to a quantum state $\rho \in \mathcal{D}(\mathcal{H})$. By using the language of probability measures we can say that every probability measure $\mathbf{P}$ on Ω can be represented by a quantum state ρ and vice versa. Thus we have the following "theorem."[9]

Theorem 4.4. (von Neumann) *Under conditions (VN1)–(VN3) (and some additional technical conditions) there is a well-defined map j: $S(\Omega) \rightarrow \mathcal{D}(\mathcal{H})$ that is one-to-one and*

$$\int_{\Omega} \xi(\omega) \mathrm{d}\mathbf{P}(\omega) = \mathrm{Tr}\ \rho\, \hat{a}, \quad \rho = j(\mathbf{P}), \hat{a} = j(\xi). \tag{4.18}$$

By using Theorem 4.4 J. von Neumann "proved"[10] the following [1].

Theorem 4.5. (Von Neumann) *Let the space of statistical states $S(\Omega)$ contain probabilities having zero dispersion. A correspondence map j between a classical statistical model*

[6] For example, in classical statistical mechanics $S(\Omega)$ is the space of *all probability measures* on phase space $\Omega = \mathbf{R}^{2n}$. In a pre-quantum classical statistical field theory that was developed in a series of works [24, 25] the space of hidden variables Ω is infinite-dimensional phase space, the space of classical fields, and the space of statistical states $S(\Omega)$ consists of Gaussian measures having very small dispersion.

[7] Different random variables from the space $V(\Omega)$ are mapped into different quantum observables (injectivity) and any quantum observable corresponds to some random variable belonging to $V(\Omega)$ (surjectivity).

[8] As J. von Neumann remarked, "the simultaneous measurability of $j(\xi_1), j(\xi_2), \ldots$ is not assumed," see [1], p. 314.

[9] The quotes are to point out that von Neumann did not proceed in a rigorous mathematical framework in this case.

[10] He did not formulate a theorem, but only an ansatz.

$$M_{\rm cl} = (S(\Omega), V(\Omega))$$

and the quantum statistical model

$$N_{\rm quant} = (\mathcal{D}(\mathcal{H}), \mathcal{L}_s(\mathcal{H}))$$

satisfying the postulates (VN1)–(VN3) (and some additional technical conditions [1]) does not exist.

4.5 Bell-type no-go theorems

As was later pointed out by several good physicists (e.g., by J. S. Bell [6] and L. Ballentine [26]), some of the von Neumann postulates of classical → quantum correspondence are non-physical. By contrast to von Neumann's theorem, in the Bell-type no-go theorem different classical quantities can correspond to the same quantum observable. It is not assumed that every self-adjoint operator corresponds to some classical quantity. It might be that some self-adjoint operators have no classical counterpart, or even physical meaning for that matter. The postulate (VN1) was deleted from the list for classical → quantum correspondence. The most doubtful postulate (VN3) was also excluded from consideration. It was not assumed that (VN2) holds.

Consider a family of spin operators,

$$\hat{\sigma}(\theta) = \cos\theta\, \hat{\sigma}_z + \sin\theta\, \hat{\sigma}_x,$$

where $\hat{\sigma}_x, \hat{\sigma}_z$ are Pauli matrices, $\theta \in [0, 2\pi)$. These operators act in the two-dimensional state space $\mathcal{H} = \mathbf{C}^2$. Let us also consider spin operators for pairs of spin-$\frac{1}{2}$ particles: $\hat{\sigma}(\theta) \otimes I$ and $I \otimes \hat{\sigma}(\theta)$. They act in the four-dimensional state space $\mathcal{H} = \mathbf{C}^2 \otimes \mathbf{C}^2$.

Bell's list of postulates on classical → quantum correspondence can be described as the following.

(B1) The image $j(V(\Omega))$ contains spectral projectors for operators $\hat{\sigma}(\theta) \otimes I$ and $I \otimes \hat{\sigma}(\theta)$ for pairs of spin-$\frac{1}{2}$ particles.

(B2) For any random variable $\xi \in V(\Omega)$, its range of values $\xi(\Omega)$ coincides with the spectrum of the operator $\hat{a} = j(\xi)$.

(B3) The image $j(S(\Omega))$ contains the singlet spin state[11]

$$\psi = \frac{1}{\sqrt{2}}(|+\rangle|-\rangle - |-\rangle|+\rangle).$$

[11] This state belongs to the four-dimensional state space $\mathcal{H} = \mathbf{C}^2 \otimes \mathbf{C}^2$. By using the tensor notations we can write $\psi = (1/\sqrt{2})\,[|+\rangle \otimes |-\rangle - |-\rangle \otimes |+\rangle]$.

Starting with any classical $\to$ quantum mapping j, we can construct a map i from quantum observables to classical random variables by setting, for $\hat{a} \in j(V(\Omega))$,

$$i(\hat{a}) = \xi_a,$$

where ξ_a belongs to the set of random variables $j^{-1}(\hat{a})$. Let us also construct a map (denoted by the same symbol i) from the space of von Neumann density operators into the space of classical probability measures by choosing a probability measure $\mathbf{P}_\rho$ belonging the set $j^{-1}(\rho)$. It should be emphasized that such maps

$$i: \mathcal{L}_s(\mathcal{H}) \to V(\Omega)$$

and

$$i: \mathcal{D}(\mathcal{H}) \to S(\Omega)$$

are not uniquely defined! Let us also note a purely mathematical problem of transition from classical $\to$ quantum to quantum $\to$ classical correspondence. Such a possibility is based on the *axiom of choice*. For example, for quantum observables we have the collection of sets $j^{-1}(\hat{a})$ of classical random variables corresponding to self-adjoint operators. We should choose from each of these sets one random variable and construct a new set – the classical image of quantum observables. The use of this axiom is not broadly accepted in the mathematical community. Finally, let us set

$$\xi_\theta = i(\hat{\sigma}(\theta) \otimes I), \qquad \xi'_\theta = i(I \otimes \hat{\sigma}(\theta)).$$

These are classical pre-images of the spin operators for pairs of spin-$\frac{1}{2}$ particles.

J. Bell also proposed use of the following postulates.

(B4) For any quantum state ρ and commuting operators $\hat{a}, \hat{b}$, the quantum and classical correlations coincide:

$$\langle \xi_a, \xi_b \rangle_{\mathbf{P}_\rho} \equiv \int_\Omega \xi_a(\omega)\xi_b(\omega) \mathrm{d}\mathbf{P}_\rho(\omega) = \langle \hat{a}\ \hat{b} \rangle_\rho \equiv \mathrm{Tr}\ \rho\ \hat{a}\hat{b}.$$

(B5) For the singlet state ψ and any θ, random variables ξ_θ and ξ'_θ are anticorrelated:

$$\xi_\theta(\omega) = -\xi'_\theta(\omega) \tag{4.19}$$

almost everywhere with respect to the probability $\mathbf{P}_\psi$.

Theorem 4.6. (Bell) *Let* $\dim \mathcal{H} = 4$. *Correspondence maps*

$$j: V(\Omega) \to \mathcal{L}_s(\mathcal{H}) \text{ and } j: S(\Omega) \to \mathcal{D}(\mathcal{H}) \tag{4.20}$$

satisfying the postulates (B1)–(B5) do not exist.

Proof. We apply Bell's inequality, Theorem 4.4, to random variables $\xi_\theta = i(\hat{\sigma}(\theta) \otimes I)$, $\xi'_\theta = i(I \otimes \hat{\sigma}(\theta))$ and to a probability measure $\mathbf{P}_\psi$ corresponding to the singlet state ψ:

$$|\langle \xi_{\theta_1}, \xi_{\theta_2} \rangle_{\mathbf{P}_\psi} - \langle \xi_{\theta_3}, \xi_{\theta_2} \rangle_{\mathbf{P}_\psi}| \leq 1 - \langle \xi_{\theta_1}, \xi_{\theta_3} \rangle_{\mathbf{P}_\psi}.$$

We remark that the postulate (B2) was used here. To prove Bell's inequality, we took into account that random variables $\xi_\theta(\omega) = \pm 1$. We now apply the anticorrelation postulate (B5) and rewrite Bell's inequality:

$$|\langle \xi_{\theta_1}, \xi'_{\theta_2} \rangle_{\mathbf{P}_\psi} - \langle \xi_{\theta_3}, \xi'_{\theta_2} \rangle_{\mathbf{P}_\psi}| \leq 1 + \langle \xi_{\theta_1}, \xi'_{\theta_3} \rangle_{\mathbf{P}_\psi}.$$

Finally, we apply the postulate (B4) and write quantum covariations, instead of classical:

$$\begin{aligned} |\mathrm{Tr}(\psi \otimes \psi)(\hat{\sigma}(\theta_1) \otimes I)(I \otimes \hat{\sigma}(\theta_2)) - \mathrm{Tr}(\psi \otimes \psi)(\hat{\sigma}(\theta_3) \otimes I)(I \otimes \hat{\sigma}(\theta_2))| \\ \leq 1 + \mathrm{Tr}(\psi \otimes \psi)(\hat{\sigma}(\theta_1) \otimes I)(I \otimes \hat{\sigma}(\theta_3)). \end{aligned}$$

But this inequality is violated for a special choice of angles $\theta_1, \theta_2, \theta_3$ – see the proof of Theorem 4.7 for details.

The postulate (B4) on the correspondence between classical and quantum correlations can be changed to a postulate about the correspondence between classical and quantum probabilities.

(W) For any quantum state ρ and commuting operators $\hat{a}, \hat{b}$, the quantum and classical joint probability distributions coincide:

$$\mathbf{P}_\rho(\omega \in \Omega\colon\ \xi_a(\omega) \in A, \xi_a(\omega) \in B) = \mathrm{Tr}\ \rho\ E^a_A E^b_B,$$

where E^a_A and E^b_B are spectral families of the operators, and A and B are arbitrary Borel subsets of the real line.

Theorem 4.7. (Wigner) *Let* $\dim \mathcal{H} = 4$. *Correspondence maps*

$$j\colon\ V(\Omega) \to \mathcal{L}_s(\mathcal{H}) \text{ and } j\colon\ S(\Omega) \to \mathcal{D}(\mathcal{H}) \tag{4.21}$$

satisfying the postulates (B1)–(B3), (W), and (B5) do not exist.

Proof. We shall apply Wigner's inequality for probabilities, Theorem 2.2:

$$\begin{aligned} \mathbf{P}_\psi(\omega \in \Omega\colon\ \xi_{\theta_1}(\omega) = +1, \xi_{\theta_2}(\omega) = +1) + \mathbf{P}_\psi(\omega \in \Omega\colon\ \xi_{\theta_2}(\omega) = -1, \xi_{\theta_3}(\omega) = +1) \\ \geq \mathbf{P}_\psi(\omega \in \Omega\colon\ \xi_{\theta_1}(\omega) = +1, \xi_{\theta_3}(\omega) = +1). \end{aligned} \tag{4.22}$$

We remark that the postulate (B2) was taken into account. We now apply the anti-correlation postulate (B5) and rewrite the Wigner inequality:

$$\mathbf{P}_\psi(\omega \in \Omega\colon\ \xi_{\theta_1}(\omega) = +1, \xi'_{\theta_2}(\omega) = -1) + \mathbf{P}_\psi(\omega \in \Omega\colon\ \xi_{\theta_2}(\omega) = -1, \xi'_{\theta_3}(\omega) = -1)$$

$$\geq \mathbf{P}_\psi(\omega \in \Omega\colon\ \xi_{\theta_1}(\omega) = +1, \xi'_{\theta_3}(\omega) = -1). \tag{4.23}$$

We apply the postulate (W) and write quantum probabilities, instead of classical ones:

$$\mathrm{Tr}(\psi \otimes \psi) E_+(\theta_1) E'_-(\theta_2) + \mathrm{Tr}(\psi \otimes \psi) E_-(\theta_2) E'_-(\theta_3)$$

$$\geq \mathrm{Tr}(\psi \otimes \psi) E_+(\theta_1) E'_-(\theta_3), \tag{4.24}$$

where $E_\pm(\theta)$ and $E'_\pm(\theta)$ are spectral projectors of the operators $\hat{\sigma}(\theta) \otimes I$ and $I \otimes \hat{\sigma}(\theta)$. We consider spin observables for pairs of spin-$\frac{1}{2}$ particles: $\sigma(\theta)$ and $\sigma'(\theta)$ corresponding to measurements on the first and second particle in a pair $s = (s_1, s_2)$. These observables are represented by operators $\hat{\sigma}(\theta) \otimes I$ and $\sigma'(\theta) \otimes I$, respectively. For the singlet state ψ we have

$$\mathbf{P}_\psi(\sigma(\theta_1) = +1, \sigma'(\theta_2) = +1) = \mathrm{Tr}(\psi \otimes \psi) E_+(\theta_1) E'_-(\theta_2) = \cos^2\left(\frac{\theta_1 - \theta_2}{2}\right),$$

$$\mathbf{P}_\psi(\sigma(\theta_3) = +1, \sigma'(\theta_2) = -1) = \mathrm{Tr}(\psi \otimes \psi) E_-(\theta_2) E'_-(\theta_3) = \sin^2\left(\frac{\theta_3 - \theta_2}{2}\right),$$

$$\mathbf{P}_\psi(\sigma(\theta_1) = +1, \sigma'(\theta_3) = +1) = \mathrm{Tr}(\psi \otimes \psi) E_+(\theta_1) E'_-(\theta_3) = \cos^2\left(\frac{\theta_1 - \theta_3}{2}\right).$$

By (4.24) we have

$$\cos^2\left(\frac{\theta_1 - \theta_2}{2}\right) + \sin^2\left(\frac{\theta_3 - \theta_2}{2}\right) \geq \cos^2\left(\frac{\theta_1 - \theta_3}{2}\right).$$

We take $\theta_1 = 0$, $\theta_2 = 6\theta$, $\theta_3 = 2\theta$ and we get the following trigonometric inequality:

$$\cos^2(3\theta) + \sin^2(2\theta) \geq \cos^2\theta.$$

It is well known [9] that this trigonometric inequality is violated for sufficiently large θ.

Considering the following weaker form of the postulate (B2),

(CHSH) For any $\xi \in V(\Omega)$,

$$\sup\{|x|\colon\ x \in \xi(\Omega)\} = \sup\{|x|\colon\ x \in \mathrm{Spectrum}(j(\xi))\},$$

we have

Theorem 4.8. *(Clauser–Horne–Shimony–Holt) Let* $\dim \mathcal{H} = 4$. *Correspondence maps*

$$j\colon\ V(\Omega) \to \mathcal{L}_s(\mathcal{H}) \text{ and } j\colon\ S(\Omega) \to \mathcal{D}(\mathcal{H}) \tag{4.25}$$

satisfying the postulates (B1), (CHSH), (B3), and (B4) do not exist.

The proof is based on the CHSH inequality. In this no-go theorem we need neither the precise anticorrelations nor the precise coincidence of ranges of values for classical variables and quantum observables. On the other hand, the postulate (B4) contains coupling between classical and quantum algebraic structures, which is not supposed in (W).

Our attitude with respect to Bell-type no-go theorems is similar to Bell's attitude with respect to other no-go theorems – the von Neumann, Jauch–Piron and Gleasons theorems ([6], pp. 4–9). As J. S. Bell did, one could speculate that some postulates about the correspondence between classical and quantum models (which were used in Bell-type no-go theorems) are non-physical. There are many things that can be questioned in Bell's arguments.

4.6 The range-of-values postulate

The proofs of Bell and Wigner no-go theorems were based on the postulate (B2) on the *coincidence of ranges of values* for classical random variables and quantum observables. Moreover, one can easily construct examples of classical random variables reproducing the EPR–Bohm correlations in the case of violation of (B2), [27].

Is the postulate (B2) really implied by the physical analysis of the situation? It seems that it is not at all! For example, Henry Stapp [11] pointed out that "The problem, basically, is that to apply quantum theory, one must divide the fundamentally undefined physical world into two idealized parts, the observed and observing system, but *the theory gives no adequate description of connection between these two parts.* The probability function is a function of degrees of freedom of the microscopic observed system, whereas the probabilities it defines are probabilities of responses of macroscopic measuring devices, and these responses are described in terms of quite different degrees of freedom." In such a situation rejection of the range-of-values condition is quite natural, since, as was pointed by Stapp, a classical random variable ξ and its quantum counterpart $\hat{a} = j(\xi)$ depend on completely different degrees of freedom. Finally, we remark that a classical model reproducing quantum probabilistic description, but violating (B2), was recently developed, see [24, 25].

The derivation of the CHSH no-go theorem was based on the CHSH postulate – the weaker form of the range-of-values coincidence postulate B2. The CHSH postulate is also a postulate about the correspondence for ranges of values for classical random variables and quantum observables, and the above arguments against (B2) can be applied against (CHSH).

Conclusion. *If the range-of-values postulates (in the forms (VN2), (B2), and (CHSH)) are rejected, then the classical probabilistic description does not contradict quantum mechanics.*

4.7 Contextuality

In this section I shall present a very general viewpoint on the role of contextuality in Bell-type no-go theorems. Bell's original viewpoint [6] on contextuality will be presented in Section 4.8. The latter contextuality we can call *simultaneous measurement contextuality* or Bell-contextuality. We reserve the term *contextuality* for our general contextuality – dependence on the whole complex of physical conditions for preparation and measurement. Although it is common in the literature to find Bell-contextuality called simply *contextuality*, using such a terminology is rather misleading because dependence on the measurements of other compatible observables is just a very special case of dependence on the general physical context.

As was rightly pointed out by Bell, the only reasonable explanation of his contextuality is *action at a distance.* Another possibility is often called *"death of reality"* [2, 3] – denying the possibility of assigning to quantum systems objective properties (such as the electron spin or the photon polarization) – but that sounds unnatural. The observation of precise (anti)correlations for the singlet state evidently contradicts the latter explanation.

In contrast to Bell-contextuality, in general, contextuality does not imply either action at a distance or "death of reality." Moreover, if one presents Bell's arguments in the general contextual approach, then in the classical (but contextual) probabilistic framework the probability of obtaining statistical data that would violate Bell's inequality equals zero, see Theorem 4.9.

4.7.1 Non-injectivity of classical → quantum correspondence

Let us first concentrate our considerations on the *classical variables → quantum observables* correspondence. As we remember, Bell, as well as Ballentine, strongly criticized von Neumann's postulate (VN1). Both Bell and Ballentine, as well as many others, emphasized that there were no physical reasons to

suppose – as von Neumann did – that for a quantum observable $\hat{a}$ its classical pre-image

$$j^{-1}(\hat{a}) = \{\xi \in V(\Omega) : \ j(\xi) = \hat{a}\}$$

should contain just one random variable. If one considers the quantum-mechanical description as an *approximative description,* then it would be quite reasonable to assume that quantum mechanics cannot distinguish sharply pre-quantum physical variables. A few different classical random variables $\xi, \eta, \ldots$ can be identified in the quantum model with the same operator $\hat{a} = j(\xi) = j(\eta) = \ldots$. Moreover, there are no reasons to hope that degeneration of the map $j\colon V(\Omega) \to \mathcal{L}_s(\mathcal{H})$ should be small. The cardinality of the set $j^{-1}(\hat{a})$ might be huge (at least for some operators).

We now consider the *classical probabilities* $\to$ *quantum states* correspondence. In the same way as for variables and observables, there are no physical reasons to assume injectivity of the map $j\colon S(\Omega) \to \mathcal{D}(\mathcal{H})$. By saying that we prepared an ensemble of systems with the fixed quantum state ρ we could not guarantee that we really prepared the fixed classical probability distribution. The set

$$j^{-1}(\rho) = \{\mathbf{P} \in S(\Omega) : \ j(\mathbf{P}) = \rho\}$$

might have huge cardinality.

Our previous considerations would induce no protest from experts in quantum foundations, and would not have for Bell either, but our following conclusions might not be appreciated so much. It should be noted that the derivations of all Bell-type no-go theorems were based on the possibility of selecting, for any quantum state (at least for the singlet state), one fixed classical probability measure $\mathbf{P}_\rho \in j^{-1}(\rho)$ and for any quantum observable (at least for spin observables) the fixed random variable $\xi \in j^{-1}(\hat{a})$.

4.7.2 Contextual opposition against Bell's approach to no-go theorems

The crucial counterargument is that at the experimental level for all Bell-type inequalities one should use data obtained in a few different runs of measurements (at least three, but in the real experimental framework four), see [8, 14, 29]. In the light of the above discussion of the non-injectivity of the classical $\to$ quantum correspondence there are no physical reasons to assume that we would be able to obtain the same classical probability distribution and the same classical random variables, for example, corresponding to spin observables.[12] We are not able to

[12] Even if we use the same macroscopic preparation and measurement devices, fluctuations of micro-parameters can induce different physical conditions [14].

guarantee that all runs of measurements are performed under the same physical conditions.

Let us consider a new random variable C describing a *complex of physical conditions* (context) during a run of measurements. Now let us try to proceed as J. Bell and his followers did by proving inequalities for correlations and probabilities. Now classical probability measures corresponding to a quantum state ρ (in particular, to the singlet state) depend on runs C: $\mathbf{P}_\rho \equiv \mathbf{P}_{\rho,C}$ as well as random variables $\xi_{a,C}(\omega), \xi_{b,C}(\omega), \xi_{c,C}(\omega)$. We start with the correlation inequality. There are three different complexes of physical conditions C_1, C_2, C_3 inducing correlations that were considered in Theorem 4.1. Here

$$\langle \xi_a, \xi_b \rangle (C_1) = \int_\Omega \xi_{a,C_1}(\omega)\xi_{b,C_1}(\omega)\mathbf{P}_{\rho,C_1}(\omega),$$

$$\langle \xi_c, \xi_b \rangle (C_2) = \int_\Omega \xi_{c,C_2}(\omega)\xi_{b,C_2}(\omega)\mathbf{P}_{\rho,C_2}(\omega).$$

If $C_1 \neq C_2$ we are not able to perform operations with integrals, which we did in Theorem 4.1. We cannot obtain Bell's inequality involving the third correlation,

$$\langle \xi_a, \xi_c \rangle (C_3) = \int_\Omega \xi_{a,C_3}(\omega)\xi_{c,C_3}(\omega)\mathbf{P}_{\rho,C_3}(\omega)$$

for a context C_3. To derive Bell's inequality, we should assume that

$$C_1 = C_2 = C_3. \tag{4.26}$$

By using the contextual framework we derived in [28, 29], we obtain generalizations of the Bell-type inequalities. Such generalized inequalities do not contradict the predictions of quantum mechanics. We also mention that a special form of contextuality (the so-called *non-reproducibility condition*) was also present in arguments of De Baere [14] against Bell's no-go theorem, see also [15]. The so-called efficiency-of-detectors (or more general unfair-sampling) argument [17] can also be considered a special form of contextuality – different contexts produce samples with different statistical properties.

The very same counterarguments can be used against the derivations of Wigner's inequality and CHSH's inequality. Let us now formalize the procedure of correspondence between classical and quantum models in the case of context-dependence. Denote by $\mathcal{C}$ the set of contexts under consideration. We suppose that on $\mathcal{C}$ there is defined a probability measure $\mathbf{Q}$. Instead of degenerate maps $j\colon V(\Omega) \to \mathcal{L}_s(\mathcal{H})$ and $j\colon S(\Omega) \to \mathcal{D}(\mathcal{H})$, we consider random maps:

$$i\colon \mathcal{C} \times \mathcal{L}_s(\mathcal{H}) \to V(\Omega),$$

$$i\colon\ \mathcal{C}\times\mathcal{D}(\mathcal{H})\to S(\Omega).$$

For any context C (considered not as a random parameter) and any quantum observable $\hat{a}$ there is uniquely defined a random variable $\xi(\omega) = i(C,\hat{a})(\omega)$ and for any quantum state ρ there is uniquely defined a probability measure $\mathbf{P} = i(C,\rho)$. In this framework we can formulate an interesting problem, namely,[13] *What is the probability of obtaining statistical data that would satisfy Bell-type inequalities?*

It is natural to assume that the probability that precisely the same complex of physical conditions would be obtained is equal to zero. Thus the probability $\mathbf{Q}$ is "*continuous*":

$$\mathbf{Q}(C) = 0 \tag{4.27}$$

for any single point $C \in \mathcal{C}$. Such a condition assumes that quantum mechanics provides only a rough description of the real physical situation.

Theorem 4.9. *Under the assumption (4.27) the probability of obtaining statistical data that would satisfy Bell-type inequalities is zero.*

Proof. Since $\mathbf{Q}(C) = 0$, the probability of obtaining in three different experiments totally identical complexes of physical conditions is zero.

4.7.3 Bell's inequality and experiment

The standard conclusion from Bell's considerations is that the experimental violation of Bell-type inequalities showed a disagreement between classical probabilistic description and experiment. Our probabilistic analysis demonstrated that such an assumption was not totally justified. In general the existence of a pre-quantum classical probabilistic model would imply only that Bell-type inequalities could hold with zero probability.

4.8 Bell-contexuality and action at a distance

In Bell's approach contexts are completely determined by compatible observables. Thus all micro-conditions of preparation and measurement are ignored. The simultaneous measurement of a compatible observable is considered as the only source of contextuality. Let $\hat{a}\in\mathcal{L}_s(\mathcal{H})$. Contexts for the a-measurement are given by all compatible observables: $C = C_b$, where $\hat{b}\in\mathcal{L}_s(\mathcal{H})$ and $[\hat{a},\hat{b}] = 0$.

[13] Thus there are two random parameters: C, describing a complex of physical conditions, and ω, describing the hidden state of a system.

In contrast to our previous considerations, C_b does not depend on a quantum state ρ or a run of preparation or measurement. Nevertheless, Bell-contextuality (as would any contextuality) also blocks the derivations of all Bell-type theorems.

As Bell pointed out, if measurements of a and b are performed in separated regions of space-time, then his contextuality can be interpreted as action at a distance:

$$i(C_b, \hat{a})(\omega) = \xi_a(C_b, \omega).$$

Thus a classical random variable $\xi_a(C_b, \omega)$ is not determined uniquely by a and it depends on the measurement of b. The latter acts instantaneously at a distance.

4.8.1 On the value of Bell's argument

In principle, Bell's argument in favor of action at a distance might stimulate investigations to find direct evidence of such an action. Unfortunately, that hasn't happened. Instead, Bell-type no-go theorems were considered as the final proofs of "quantum non-locality." However, as we have seen, Bell-contextuality and consequent non-locality is only one of many possibilities for blocking the derivations of no-go theorems. One could not draw any definite conclusion from Bell-type theorems besides the evident remark that, since we do not know a pre-quantum classical statistical model and the rules of classical $\rightarrow$ quantum correspondence, we can play a lot with such hypothetical rules. The known no-go theorems are the result of such games with correspondence rules.

Personally I think that the violation of the postulate regarding coincidence of ranges of values or (and) general preparation and measurement contextuality provide essentially more natural possibilities for blocking the Bell-type theorems than does non-locality. My conclusion is the following. *The main value of Bell's arguments was the great stimulation of experimental technologies for working with entangled photons.*

Acknowledgments

This work was supported in part by the EU Human Potential Programme, contract HPRN–CT–2002–00279 (Network on Quantum Probability and Applications), and Profile Mathematical Modeling in Physics and Cognitive Science of Växjö University.

References

[1] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton, NJ: Princeton University Press, 1955).

[2] B. d'Espagnat, *Veiled Reality. An Anlysis of Present-day Quantum Mechanical Concepts* (New York: Addison-Wesley, 1995).

[3] A. Shimony, *Search for a Naturalistic World View* (Cambridge: Cambridge University Press, 1993).

[4] A. Yu. Khrennikov (ed.), *Foundations of Probability and Physics* (Singapore: World Scientific, 2001); *Quantum Theory: Reconsideration of Foundations* (Växjö: Växjö University Press, 2002); *Foundations of Probability and Physics – 2* (Växjö: Växjö University Press, 2003); *Foundations of Probability and Physics – 3* (Melville, NY: American Institute of Physics, 2005).

[5] A. Einstein, B. Podolsky, and N. Rosen, *Phys. Rev.* **47**, 777–780 (1935).

[6] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge: Cambridge University Press, 1987).

[7] J. S. Bell, *Rev. Mod. Phys.* **38**, 447–452 (1966).

[8] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Phys. Rev. Lett.* **49**, 1804–1806 (1969); J. F. Clauser and A. Shimony, *Rep. Progr. Phys.* **41** 1881–1901 (1978).

[9] E. P. Wigner, On hidden variables and quantum mechanical probabilities, *Am. J. Phys.* **38**, 1005 (1970).

[10] A. Aspect, J. Dalibard, and G. Roger, *Phys. Rev. Lett.* **49**, 1804–1807 (1982); D. Home and F. Selleri, *Nuovo Cim. Riv.,* **14**, 2–176 (1991); P. H. Eberhard, *Nuovo Cim. B* **38**, 75–80 (1977); *Phys. Rev. Lett.* **49**, 1474–1477 (1982); A. Peres, *Am. J. Phys.* **46**, 745–750 (1978); P. H. Eberhard, *Nuovo Cim. B*, **46**, 392–419 (1978); J. Jarrett, *Noûs* **18**, 569 (1984); I. Pitowsky, *Phys. Rev. Lett.* **48**, 1299–1302 (1982); *Phys. Rev. D* **27**, 2316–2326 (1983); S. P. Gudder, *J. Math. Phys.* **25**, 2397–2401 (1984).

[11] H. P. Stapp, *Phys. Rev. D* **3**, 1303–1320 (1971).

[12] L. Accardi, *Urne e camaleoni: dialogo sulla realtà, le leggi del caso e la teoria quantistica* (Rome: Il Saggiatore, 1997); The probabilistic roots of the quantum mechanical paradoxes, in *The Wave–Particle Dualism. A Tribute to Louis de Broglie on His 90th Birthday*, S. Diner, D. Fargue, G. Lochak, and F. Selleri (eds.) (Dordrecht: Reidel, 1970), pp. 45–55.

[13] A. Fine, *Phys. Rev. Lett.* **48**, 291–295 (1982); P. Rastal, *Foundations Phys.* **13**, 555 (1983).

[14] W. De Baere, *Lett. Nuovo Cim.* **39**, 234–238 (1984); S. P. Gudder, *J. Math. Phys.* **25**, 2397–2401 (1984).

[15] W. De Muynck and W. De Baere, *Ann. Israel Phys. Soc.* **12**, 1–22 (1996); W. De Muynck, W. De Baere, and H. Marten, *Foundations Phys.* **24**, 1589–1663 (1994); W. De Muynck and J. T. Stekelenborg, *Ann. Phys.* **45**(7), 222–234 (1988).

[16] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, *Phys. Rev. Lett.* **81**, 23, 5039 (1998).

[17] P. Pearle, *Phys. Rev. D* **2**, 1418, (1970); N. Gisin and B. Gisin, *Phys. Lett. A* **260**, 323 (1999); Jan-Åke Larsson, *Quantum Paradoxes, Probability Theory, and Change of Ensemble* (Linköping: Linköping University Press, 2000); G. Adenier and A. Khrennikov, Anomalies in EPR-Bell experiments, in *Quantum Theory: Reconsideration of Foundations – 3* (Melville, NY: American Institute of Physics 2006), pp. 283–293.

[18] C. H. Thompson and H. Holstein, quant-ph/0210150 (2002); A. F. Kracklauer, quant-ph/0302113 (2003); B. C. Sanctuary, quant-ph/0304186 (2003).
[19] W. Muckenheim, *Phys. Rep.* **133**, 338–401 (1986); A. Yu. Khrennikov, *Dokl. Akad. Nauk SSSR, Ser. Matem.* **322**, 1075–1079 (1992); *J. Math. Phys.* **32**, 932–937 (1991); *Phys. Lett. A* **200**, 119–223 (1995); *Physica A* **215**, 577–587 (1995); *Int. J. Theor. Phys.* **34**, 2423–2434 (1995); *J. Math. Phys.* **36**, 6625–6632 (1995); A. Yu. Khrennikov, *p-Adic Valued Distributions in Mathematical Physics* (Dordrecht: Kluwer, 1994); A. Yu. Khrennikov, *Non-Archimedean Analysis: Quantum Paradoxes, Dynamical Systems and Biological Models* (Dordrecht: Kluwer, 1997).
[20] A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsreich* (Berlin: Springer, 1933; English translation as *Foundations of the Probability Theory* (New York: Chelsea, 1956).
[21] A. N. Kolmogorov, The theory of probability, in *Mathematics, Its Content, Methods, and Meaning* A. D. Alexandrov, A. N. Kolmogorov, and M. A. Lavrent'ev (eds.) (Boston, MA: MIT Press, 1965), pp. 110–118.
[22] B. V. Gnedenko, *The Theory of Probability* (New York: Chelsea, 1962).
[23] A. N. Shiryayev, *Probability* (New York: Springer, 1991).
[24] A. Yu. Khrennikov, *J. Phys. A* **38**, 9051–9073 (2005); *Foundations Phys. Lett.* **18**, 637–650 (2006); *Phys. Lett. A* **357**, 171–176 (2006).
[25] A. Yu. Khrennikov, Quantum mechanics as an asymptotic projection of statistical mechanics of classical fields, in *Quantum Theory: Reconsideration of Foundations – 3*, G. Adenier, A. Yu. Khrennikov, and Th. M. Nieuwenhuizen (eds.) (Melville, NY: American Institute of Physics, 2006), pp. 179–197.
[26] L. E. Ballentine, *Rev. Mod. Phys.* **42**, 358–381 (1970).
[27] A. Yu. Khrennikov and I. V. Volovich, *Soft Computing* **10**, 521–529 (2005).
[28] A. Yu. Khrennikov, *Nuovo Cim. B* **115**, 179–184 (1999); *J. Math. Phys.* **41**, 5934–5944 (2000); *J. Math. Phys.* **41**, 1768–1777 (2000); *Phys. Lett. A* **278**, 307–314 (2001).
[29] A. Yu. Khrennikov, *Interpretations of Probability*, 2nd edn. (Utrecht: VSP, 2004).

5

Probabilistic theories: What is special about Quantum Mechanics?

Giacomo Mauro D'Ariano

To my friend and mentor, Professor Attilio Rigamonti.

Unperformed experiments have no results.
Asher Peres

5.1 Introduction

More than a century after its birth, quantum mechanics (QM) remains mysterious. We still don't have general principles from which to derive its remarkable mathematical framework, as happened for the amazing Lorentz transformations, which were rederived by Einstein from the invariance of physical laws in inertial frames and from the constancy of the speed of light.

Despite the utmost relevance of the problem of deriving QM from operational principles, research efforts in this direction have been sporadic. The deepest of the early attacks on the problem were the works of Birkhoff, von Neumann, Jordan, and Wigner, attempting to derive QM from a set of axioms with as much physical significance as possible [1, 2]. The general idea in Ref. [1] is to regard QM as a new kind of *prepositional calculus*, a proposal that spawned the research line of *quantum logic*, which is based on von Neumann's observation that the two-valued observables – represented in his formulation of QM by orthogonal projection operators – constitute a kind of "logic" of experimental propositions. After a hiatus of two decades of neglect, interest in quantum logic was revived by Varadarajan [3], and most notably by Mackey [4], who axiomatized QM within an operational framework, with the single exception of an admittedly *ad hoc* postulate, which represents the propositional calculus mathematically in the form of an orthomodular lattice. The most significant extension of Mackey's work is the general representation theorem of Piron [5].

Philosophy of Quantum Information and Entanglement, ed. A. Bokulich and G. Jaeger. Published by Cambridge University Press.

In the early work [2], Jordan, von Neumann, and Wigner considered the possibility of a commutative algebra of observables, with a product that needs only to define squares and sums of observables – the so-called *Jordan product* of observables a and b: $a \circ b := (a+b)^2 - a^2 - b^2$. However, such a product is generally non-associative and non-distributive with respect to the sum, and the quantum formalism follows only with additional axioms with no clear physical significance – e.g., a distributivity axiom for the Jordan product. Segal [6] later constructed an (almost) fully operational framework (with no experimental definition of the sum of observables) that allows generally non-distributive algebras of observables, but with a resulting mathematical framework largely more general than QM. As a result of this line of investigation, the purely algebraic formulation of QM gained in popularity versus the original Hilbert-space axiomatization.

In the algebraic axiomatization of QM, a physical system is defined by its C^*-algebra of observables (with identity), and the states of the system are identified with normalized positive linear functionals over the algebra, corresponding to the probability rules of measurements of observables. Indeed, the C^*-algebra of observables is more general than QM, since it includes classical mechanics as a special case, and generally describes any quantum–classical hybrid, thus being equivalent to QM with super-selection rules. Since in practice two observables are not distinguishable if they always exhibit the same probability distributions, at the operational level one can always take the set of states as *observable-separating* – in the sense that there are no different observables having the same probability distribution for all states. Conversely the set of observables is *state-separating*, i.e., there are no different states corresponding to the same probability distribution for all observables. Notice that, in principle, there exist different observables with the same expectation for all states, but higher moments will be different.[1]

The algebra of observables is generally considered to be more "operational" than the usual Hilbert-space axiomatization; however, little more is gained than a representation-independent mathematical framework. Indeed, the algebraic framework is unable to provide operational rules for how to measure sums and products of non-commuting observables.[2] The sum of two observables cannot be given an operational meaning, since a procedure involving the measurements of the two addenda would unavoidably assume that their respective measurements are jointly executable on the same system – i.e., the observables are *compatible*. The same

[1] This is not the case when one considers only *sharp observables*, for which there always exists a state such that the expectation of any function of the observable equals the function of the expectation. However, operationally we cannot rely on such a concept to define the general notion of an observable, since we cannot reasonably assume its feasibility (actual measurements are non-sharp).

[2] The spectrum of the sum is generally different from the sum of the spectra of the addenda, e.g., the spectra of xp_y and yp_x are both $\mathbb{R}$, whereas the angular-momentum component $xp_y - yp_x$ has a discrete spectrum. The same is true for the product.

reasoning holds for the product of two observables. A sum-observable defined as the one having expectation equal to the sum of expectations for all states [7] is clearly not unique, due to the existence of observables having the same expectation for all states, but with different higher moments. The only well-defined procedures are those involving single observables, such as the measurement of a *function of a single observable*, which operationally consists in just taking the function of the outcome.

The Jordan symmetric product has been regarded as a great advance in view of an operational axiomatization, since, in addition to being Hermitian (observables are Hermitian), it is defined only in terms of squares and sums of observables – i.e., without products. The definition of $a \circ b$, however, still relies on the notion of a sum of observables, which has no operational meaning. Remarkably, Alfsen and Shultz [8, 9] succeeded in deriving the Jordan product from solely geometrical properties of the convex set of states – e.g., orientability and faces shaped as Euclidean balls – however, again with no operational meaning. The problem with the Jordan product is that, in addition to not necessarily being associative, it is not even distributive, as the reader can easily check. It turns out that, modulo a few topological assumptions, the Jordan algebras can be embedded in the algebra $\mathsf{Lin}(\mathsf{H})$ of operators over the Hilbert space H, whereby $a \circ b = ab + ba$. Such assumptions, however, are still not operational. For a further critical overview of these earlier attempts at an operational axiomatization of QM, the reader is also directed to the recent books of Strocchi [7] and Thirring [10].

After a long suspension of research on the axiomatic approach – notably interrupted by the work of Ludwig and his school [11] – in the last few years the new field of quantum information has renewed interest in the problem of operational axiomatization of QM, having been boosted by the new experience on multipartite systems and *entanglement*. In his seminal paper [12] Hardy derived QM from five "reasonable axioms," which, more than being truly operational, are motivated on the basis of simplicity and continuity criteria, with the assumption of a finite number of perfectly discriminable states. His axiom 4, however, is still purely mathematical, and is directly related to the tensor-product rule for composite systems. In another popular paper [13], Clifton, Bub, and Halvorson have shown how three fundamental information-theoretic constraints – (a) the no-signaling constraint, (b) the no-broadcasting constraint, and (c) the impossibility of unconditionally secure bit commitment – suffice to entail that the observables and state space of a physical theory are quantum mechanical. Unfortunately, the authors started from a C^*-algebraic framework for observables, which, as already discussed, has little operational basis, and already coincides with the quantum–classical hybrid. Therefore, more than deriving QM, their informational principles just force the C^*-algebra of observables to be non-Abelian.

In Ref. [14][3] I showed how it is possible to derive the formulation of QM in terms of observables represented as Hermitian operators over Hilbert spaces with the right dimensions for the tensor product, starting from a few operational axioms. However, it is not clear yet whether such a framework is sufficient to identify QM (or the quantum–classical hybrid) as the only probabilistic theory resulting from axioms. Later, in Refs. [17–19], I showed how a C^*-algebraic framework for transformations (not for observables!) naturally follows from an operational probabilistic framework.

A very recent and promising direction for attacking the problem of QM axiomatization consists in positioning QM within the landscape of general probabilistic theories, including theories with non-local correlations stronger than the quantum ones, e.g., for the Popescu–Rohrlich boxes (PR boxes) [20]. Such theories have correlations that are "stronger" than the quantum ones – in the sense that they violate the quantum Cirel'son bound [21] – although they are still non-signaling, thus revealing the fortuitousness of the peaceful coexistence of QM and special relativity, in contrast with the claimed implication of QM linearity from the no-signaling condition [22]. Within the framework of the PR boxes general versions of the no-cloning and no-broadcasting theorems have been proved [23]. In Ref. [24] it has been shown that certain features generally thought of as specifically quantum are indeed present in all except classical theories. These include the non-unique decomposition of a mixed state into pure states, disturbance on measurement (related to the possibility of secure key distribution), and the no-cloning constraint. More recently, necessary and sufficient conditions have been established for teleportation [25], i.e., for reconstructing the state of a system on a remote identical system, using only local operations and joint states. In all these works quantum information has inspired the consideration of task-oriented axioms in a general operational framework that can incorporate QM, classical theory, and other non-signaling probabilistic theories (for an illustration of this general point of view see also Ref. [26]).

In this chapter I will consider the possibility of deriving QM as the mathematical representation of a *fair operational framework*, i.e., a set of rules that allows the experimenter to make predictions regarding future *events* on the basis of suitable *tests*, in a spirit close to Ludwig's axiomatization [11]. *States* are simply the compendia of probabilities for all possible outcomes of any test. I will consider a very general class of probabilistic theories, and examine the consequences of two postulates that need to be satisfied by any fair operational framework:

NSF: *no signaling from the future*, implying that it is possible to make predictions based on present tests;

[3] Most of the results of Ref. [14] were originally conjectured in Refs. [15] and [16].

PFAITH: *existence of preparationally faithful states*, implying the possibility of preparing any state and calibrating any test.

NSF is implicit in the very definition of conditional probabilities for cascade tests, entailing that *events are identified with transformations*, whence *evolution is identified with conditioning*. As we will see, such identifications lead to the notion of *effect* of Ludwig, i.e., the equivalence class of events occurring with the same probability for all states. I will show how we can introduce operationally a linear-space structure for effects. I will then show how all theories satisfying NSF admit a C^*-algebra representation of events as linear transformations of effects.

On the basis of a very general notion of dynamical independence, entailing the definition of a *marginal state*, it is immediately seen that all these theories are *non-signaling*, which is the current way of saying that the theories satisfy the principle of *Einstein locality*, namely that there can be no detectable effect on a system due to anything done to another non-interacting system. This is clearly another requirement for a fair operational framework. Postulate PFAITH then implies the *local observability principle*, namely the possibility of achieving an informationally complete test using only local tests – another requirement for a fair operational framework. The same postulate also implies many other features that are typically quantum, such as the tensor-product structure for the linear spaces of states and effects, the isomorphism of cones of states and effects (a weaker version of quantum self-duality), the so-called EPR cheating in bit commitment (which in Ref. [13], we remind the reader, was itself used as a postulate to derive QM), and many more. Dual to Postulate PFAITH an analogous postulate for effects would give additional quantum features, such as teleportation. However, all possible consequences of these postulates still need to be investigated, and it is not clear yet whether one can derive QM from these principles only.

In order to provide a route for seeking new candidates for operational postulates one can short-circuit the axiomatic framework to select QM using a mathematical postulate dictated by what is really special about the quantum theory, namely that not only transformations but also *effects form a C^*-algebra* (more precisely, this is true for all hybrid quantum–classical theories, i.e., those corresponding to QM plus super-selection rules). However, whereas the sum of effects can be operationally defined, their composition has no operational meaning, since the notion itself of "effect" abhors any kind of composition. I will then show that with another natural postulate,

AE: atomicity of evolution,

together with the mathematical postulate

CJ: Choi–Jamiolkowski isomorphism [27, 28],

it is possible to identify effects with "atomic" events, i.e., elementary events that cannot be refined as the union of events. Via the composition of atomic events we can then define the composition of effects, thus selecting the quantum–classical hybrid among all possible general probabilistic theories (including the PR boxes, which indeed satisfy both NSF and PFAITH).

The CJ isomorphism looks natural in an operational context, and it is hoped that it will be converted soon into an operational postulate.

The present operational axiomatization will adhere to the following three general principles:

(1) **(Strongly Copenhagen)** Everything is defined operationally, including all mathematical objects. Operationally indistinguishable entities are identified.
(2) **(Mathematical closure)** Mathematical completion is taken for convenience.
(3) **(Operational closure)** Every operational option that is implicit in the formulation is incorporated in the axiomatic framework.

An example of the application of the strongly-Copenhagen principle is the notion of *system*, which here I will identify with a collection of tests – the tests that can be performed over the system. A typical case of operational identification is that of events occurring with the same probability and producing the same conditioning. Another case is the statement that the set of states is separating for effects and vice versa. Examples of mathematical closure are the norm closure, the algebraic closure, and the linear span. It is unquestionable that these are always idealizations of operational limiting cases, or they are introduced just to simplify the mathematical formulation (e.g., real numbers versus the "operational" rational numbers). Operational completeness, on the other hand, does not affect the corresponding mathematical representation, since every incorporated option is already implicit in the formulation. This is the case, for example, for convex closure, closure under coarse-graining, etc., which are already implicit in the probabilistic formulation.

5.2 C^*-Algebra representation of probabilistic theories

5.2.1 Tests and states

A probabilistic operational framework is a collection of **tests**[4] $\mathbb{A}, \mathbb{B}, \mathbb{C}, \ldots$ each being a complete collection $\mathbb{A} = \{\mathscr{A}_i\}$, $\mathbb{B} = \{\mathscr{B}_j\}$, $\mathbb{C} = \{\mathscr{C}_k\}$, ...of mutually

[4] The present notion of test corresponds to that of **experiment** of Ref. [14]. Quoted from that reference: "An experiment on an object system consists in making it interact with an apparatus, which will produce one of a set of possible events, each one occurring with some probability. The probabilistic setting is dictated by the need of experimenting with partial *a priori* knowledge about the system (and the apparatus). In the logic of performing experiments to predict results of forthcoming experiments in similar preparations, the information gathered in an experiment will concern whatever kind of information is needed to make predictions, and this, by definition, is the *state* of the object system at the beginning of the experiment. Such information is gained from the knowledge of which transformation occurred, which is the 'outcome' signaled by the apparatus."

exclusive **events** $\mathscr{A}_i, \mathscr{B}_j, \mathscr{C}_k, \ldots$ occurring probabilistically[5]; events that are mutually exclusive are often called **outcomes**. The same event can occur in different tests, with occurrence probability independent of the test. A **singleton test** – also called a **channel** – $\mathbb{D} = \{\mathscr{D}\}$ is **deterministic**: it represents a non-test, i.e., a free evolution. The **union** $\mathscr{A} \cup \mathscr{B}$ of two events corresponds to the event in which either $\mathscr{A}$ or $\mathscr{B}$ occurred, but it is unknown which one. A **refinement** of an event $\mathscr{A}$ is a set of events $\{\mathscr{A}_i\}$ occurring in some test such that $\mathscr{A} = \cup_i \mathscr{A}_i$. The experiment $\mathbb{A}$ itself can be regarded as the deterministic event corresponding to the complete union of its outcomes, and when regarded as an event it will be denoted by the different notation $\mathscr{D}_{\mathbb{A}}$. The opposite event of $\mathscr{A}$ in $\mathbb{A}$ will be denoted as $\overline{\mathscr{A}} := \complement_{\mathbb{A}}\mathscr{A}$.[6] The union of events transforms a test $\mathbb{A}$ into a new test $\mathbb{A}'$, which is a **coarse-graining** of $\mathbb{A}$, e.g., $\mathbb{A} = \{\mathscr{A}_1, \mathscr{A}_2, \mathscr{A}_3\}$ and $\mathbb{A}' = \{\mathscr{A}_1, \mathscr{A}_2 \cup \mathscr{A}_3\}$. Vice versa, we will call $\mathbb{A}$ a **refinement** of $\mathbb{A}'$.

The **state** ω describing the preparation of the system is the probability rule $\omega(\mathscr{A})$ for any event $\mathscr{A} \in \mathbb{A}$ occurring in any possible test $\mathbb{A}$.[7] For each test $\mathbb{A}$ we have the completeness $\sum_{\mathscr{A}_j \in \mathbb{A}} \omega(\mathscr{A}_j) = 1$. States themselves are considered as special tests: the **state-preparations**.

5.2.2 Cascading, conditioning, and transformations

The **cascade** $\mathbb{B} \circ \mathbb{A}$ of two tests $\mathbb{A} = \{\mathscr{A}_i\}$ and $\mathbb{B} = \{\mathscr{B}_j\}$ is the new test with events $\mathbb{B} \circ \mathbb{A} = \{\mathscr{B}_j \circ \mathscr{A}_i\}$, where $\mathscr{B} \circ \mathscr{A}$ denotes the **composite event** $\mathscr{A}$ "followed by" $\mathscr{B}$ satisfying the following

Postulate NSF (No signaling from the future). *The marginal probability* $\sum_{\mathscr{B}_j \in \mathbb{B}} \omega(\mathscr{B}_j \circ \mathscr{A})$ *of any event* $\mathscr{A}$ *is independent of test* $\mathbb{B}$*, and is equal to the probability with no test* $\mathbb{B}$*, namely*

$$\sum_{\mathscr{B}_j \in \mathbb{B}} \omega(\mathscr{B}_j \circ \mathscr{A}) =: f(\mathbb{B}, \mathscr{A}) \equiv \omega(\mathscr{A}), \quad \forall \mathbb{B}, \mathscr{A}, \omega. \tag{5.1}$$

[5] Also A. Rényi [29] calls our test "experiment." More precisely, he defines an experiment $\mathbb{A}$ as the pair $\mathbb{A} = (\mathfrak{X}, \mathcal{A})$ made of the *basic space* $\mathfrak{X}$ – the collection of outcomes – and of the σ-algebra of events $\mathcal{A}$. Here, to decrease the mathematical load of the framework, we conveniently identify the experiment with the basic space only, and consider a different σ-algebra (e.g., a coarse-graining) as a new test made of new mutually exclusive events. Indeed, since we are considering only discrete basic spaces, we can put basic space and σ-algebra in one-to-one correspondence, by taking $\mathcal{A} = 2^{\mathfrak{X}}$ – the power set of $\mathfrak{X}$ – and, vice versa, $\mathfrak{X}$ as the collection of the minimal intersections of elements of $\mathcal{A}$.

[6] By adding the intersection of events, one builds up the full *Boolean algebra of events* (see, e.g., Ref. [29]).

[7] By definition the state is the collection of the variables of a system knowledge of which is sufficient to make predictions. In the present context, it allows one to predict the results of tests, whence it is the probability rule for all events in any conceivable test.

NSF is part of the very definition of test-cascade; however, we treat it as a separate postulate, since it corresponds to the **choice of the arrow of time**.[8] The interpretation of the test-cascade $\mathbb{B} \circ \mathbb{A}$ is that "test $\mathbb{A}$ can influence test $\mathbb{B}$ but not vice versa."[9] Postulate NSF allows one to define the conditioned probability $p(\mathscr{B}|\mathscr{A}) = \omega(\mathscr{B} \circ \mathscr{A})/\omega(\mathscr{A})$ of event $\mathscr{B}$ occurring conditionally on the previous occurrence of event $\mathscr{A}$. It also guarantees that the probability of $\mathscr{B}$ remains independent of the test $\mathbb{B}$ when conditioned.

Conditioning sets a new probability rule corresponding to the notion of a **conditional state** $\omega_{\mathscr{A}}$, which gives the probability that an event occurs, knowing that event $\mathscr{A}$ has occurred with the system prepared in the state ω, namely $\omega_{\mathscr{A}} \doteq \omega(\cdot \circ \mathscr{A})/\omega(\mathscr{A})$.[10] We can now regard the event $\mathscr{A}$ as transforming with probability $\omega(\mathscr{A})$ the state ω to the (unnormalized) state[11] $\mathscr{A}\omega$ given by

$$\mathscr{A}\omega := \omega(\cdot \circ \mathscr{A}). \tag{5.2}$$

Therefore, the notion of cascade and postulate NSF entail the identification

$$\textbf{event} \equiv \textbf{transformation},$$

which in turn implies the equivalence[12]

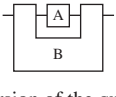
$$\textbf{evolution} \equiv \textbf{state-conditioning}.$$

Notice that operationally a transformation $\mathscr{A}$ is completely specified by all the joint probabilities in which it is involved, whence it is unequivocally given by the probability rule $\mathscr{A}\omega = \omega(\cdot \circ \mathscr{A})$ for all states ω. This is equivalent to specifying both the conditional state $\omega_{\mathscr{A}}$ and the probability $\omega(\mathscr{A})$ for all possible states ω, due to the identity

$$\mathscr{A}\omega = \omega(\mathscr{A})\omega_{\mathscr{A}}. \tag{5.3}$$

[8] Postulate NSF is not just a Kolmogorov consistency condition for marginals of a joint probability. In fact, even though the marginal over test $\mathbb{B}$ in (5.1) is obviously the probability of $\mathscr{A}$, such probability in principle depends on the test $\mathbb{B}$, since the joint probability generally depends on it. Indeed, the marginal over entry $\mathscr{A}$ does generally depend on the past test $\mathbb{A} \ni \mathscr{A}$. Such asymmetry of the joint probability under marginalization over future or past tests represents *the choice of the arrow of time*. Of course one could have assumed the opposite postulate of no signaling from the past, considering conditioning from the future instead, thus reversing the arrow of time. Postulate NSF introduces conditioning from tests, and is part of the very definition of temporal cascade-tests. The need to consider NSF as a postulate was noticed for the first time by Masanao Ozawa (private communication).

[9] One could also define more general cascades not in time, e.g., the circuit diagram.



This would have given rise to a probabilistic version of the quantum comb theory of Ref. [30].

[10] Throughout, the central dot "·" denotes the location of the pertinent variable.

[11] This is the same as the notion of *quantum operation* in QM, which gives the conditioning $\omega_{\mathscr{A}} = \mathscr{A}\omega/(\mathscr{A}\omega(\mathscr{I}))$, or, in other words, the analogue of the quantum Schrödinger-picture evolution of states.

[12] Clearly this includes the deterministic singleton-tests $\mathbb{D} = \{\mathscr{D}\}$ – the analogs of quantum channels, including unitary evolutions.

In particular the **identity transformation** $\mathscr{I}$ is completely specified by the rule $\mathscr{I}\omega = \omega$ for all states ω.

5.2.3 Systems

In a pure Copenhagen spirit we will identify a **system** S with a collection of **tests** $\mathsf{S} = \{\mathbb{A}, \mathbb{B}, \mathbb{C}, \ldots\}$, the collection being operationally closed under coarse-graining, convex combination, conditioning, and cascading, and will include all states as special tests. Closure under cascading is equivalent to considering **mono-systemic evolution**, i.e., in which there are only tests for which the output system is the same as the input one.[13] The operator has always the option of performing repeated tests, together with (randomly) alternating tests – say $\mathbb{A}$ and $\mathbb{B}$ – in different proportions – say p and $1 - p$ $(0 < p < 1)$ – thus achieving the test $\mathbb{C}_p = p\mathbb{A} + (1 - p)\mathbb{B}$ which is the **convex combination** of tests $\mathbb{A}$ and $\mathbb{B}$, and is given by $\mathbb{C}_p = \{p\mathscr{A}_1, p\mathscr{A}_2, \ldots, (1 - p)\mathscr{B}_1, (1 - p)\mathscr{B}_2, \ldots\}$, where $p\mathscr{A}$ is the same event as $\mathscr{A}$, but occurring with a probability rescaled by p. Since we will consider always closure under all the operator's options (this is our *operational closure*), we will take the system also to be closed under such convex combination. In particular, the set of all states of the system[14] is closed under convex combinations and under conditioning, and we will denote by $\mathfrak{S}(\mathsf{S})$ ($\mathfrak{S}$ for short) the convex set of all possible states of system S. We will often use the colloquialism "for all possible states ω" meaning $\forall\omega \in \mathfrak{S}(\mathsf{S})$, and we will do similarly for other operational objects.

In the following we will denote the set of all possible transformations/events by $\mathfrak{T}(\mathsf{S})$, $\mathfrak{T}$ for short. The convex structure of S entails a convex structure for $\mathfrak{T}$, whereas the cascade of tests entails the composition of transformations. The latter, together with the existence of the identity transformation $\mathscr{I}$, gives to $\mathfrak{T}$ the structure of a *convex monoid*.

5.2.4 Effects

From the notion of a conditional state two complementary types of equivalences for transformations follow: the *conditional* and the *probabilistic* equivalence. The transformations $\mathscr{A}_1$ and $\mathscr{A}_2$ are **conditioning-equivalent** when $\omega_{\mathscr{A}_1} = \omega_{\mathscr{A}_2} \forall\omega \in \mathfrak{S}$,

[13] We could have considered more generally tests in which the output system is different from the input one, in which case the system is no longer closed under a test-cascade, and, instead, there are cascades of tests from different systems. This would give more flexibility to the axiomatic approach, and could be useful for proving some theorems related to multipartite systems made of different systems. The fact that there are different systems would impose constraints on the cascades of tests, corresponding to allowing only some particular words made of the "alphabet" $\mathbb{A}, \mathbb{B}, \ldots$ of tests, and the system would then correspond to a "language" (see Ref. [31] for a similar framework). Such generalization will be thoroughly analyzed in a forthcoming publication.

[14] At this stage such a set does not necessarily contain all *in-principle* possible states. The extension will be done later, after defining effects.

namely when they produce the same conditional state for all prior states ω. On the other hand, the transformations $\mathscr{A}_1$ and $\mathscr{A}_2$ are **probabilistically equivalent** when $\omega(\mathscr{A}_1) = \omega(\mathscr{A}_2)\ \forall\omega \in \mathfrak{S}$, namely when they occur with the same probability.[15] Since operationally a transformation $\mathscr{A}$ is completely specified by the probability rule $\mathscr{A}\omega$ for all states, it follows that two transformations $\mathscr{A}_1$ and $\mathscr{A}_2$ are fully **equivalent** (i.e., operationally indistinguishable) when $\mathscr{A}_1\omega = \mathscr{A}_2\omega$ for all states ω. We will identify two equivalent transformations, and denote the equivalence simply as $\mathscr{A}_1 = \mathscr{A}_2$. From identity (5.3) it follows that two transformations are equivalent if and only if they are both conditioning and probabilistically equivalent.

A probabilistic equivalence class of transformations defines an **effect**.[16] In the following we will denote effects with lower-case letters $a, b, c, \ldots$ and denote by $[\mathscr{A}]_{\text{eff}}$ the effect containing transformation $\mathscr{A}$. We will also write $\mathscr{A} \in a$ meaning that "the transformation $\mathscr{A}$ belongs to the equivalence class a," or "$\mathscr{A}$ has effect a," and write "$\mathscr{A} \in [\mathscr{B}]_{\text{eff}}$" to say that "$\mathscr{A}$ is probabilistically equivalent to $\mathscr{B}$." Since by definition $\omega(\mathscr{A}) = \omega([\mathscr{A}]_{\text{eff}})$, hereafter we will legitimately write the variable of the state as an effect, e.g., $\omega(a)$. The **deterministic effect** will be denoted by e, corresponding to $\omega(e) = 1$ for all states ω. We will denote the set of effects for a system S as $\mathfrak{E}(\mathsf{S})$, or just $\mathfrak{E}$ for short. The set of effects inherits a convex structure from that of transformations.

By the same definition of state – as probability rule for transformations – states are separated by effects (whence also by transformations[17]), and, conversely, effects are separated by states. Transformations are separated by states in the sense that $\mathscr{A} \neq \mathscr{B}$ iff $\mathscr{A}\omega \neq \mathscr{B}\omega$ for some state. As a consequence, it may happen that the introduction of a new state via some new preparation (such as introducing additional systems) will separate two previously indiscriminable transformations, in which case we will include the new state (and all convex combinations with it) in $\mathfrak{S}(\mathsf{S})$, and we will complete the system S accordingly. We will end with $\mathfrak{S}(\mathsf{S})$ separating $\mathfrak{T}(\mathsf{S})$ and $\mathfrak{E}(\mathsf{S})$, and $\mathfrak{E}(\mathsf{S})$ separating $\mathfrak{S}(\mathsf{S})$.

The identity $\omega_{\mathscr{A}}(\mathscr{B}) \equiv \omega_{\mathscr{A}}([\mathscr{B}]_{\text{eff}})$ implies that $\omega(\mathscr{B} \circ \mathscr{A}) = \omega([\mathscr{B}]_{\text{eff}} \circ \mathscr{A})$ for all states ω, leading to the chaining rule $[\mathscr{B}]_{\text{eff}} \circ \mathscr{A} = [\mathscr{B} \circ \mathscr{A}]_{\text{eff}}$, corresponding to the "Heisenberg-picture" evolution in terms of transformations acting on effects.

[15] In the papers [14–17] I called the conditional equivalence *dynamical equivalence*, since the two transformations will effect the same state change. However, one should more properly regard the "dynamical" change of the state ω due to the transformation $\mathscr{A}$ as the unnormalized state $\mathscr{A}\omega$, but the two transformations $\mathscr{A}$ and $\mathscr{B}$ will be fully equivalent when $\mathscr{A}\omega = \mathscr{B}\omega$ for all states ω. Moreover, in the same papers I called the probabilistic equivalence *informational equivalence*, since the two transformations will give the same information about the state. The new nomenclature has a more immediate meaning.

[16] This is the same notion of "effect" introduced by Ludwig [11].

[17] In fact, $\mathscr{A}\omega \neq \mathscr{A}\zeta$ for $\mathscr{A} \in \mathfrak{T}$ means that there exists an effect c such that $\mathscr{A}\omega(c) \neq \mathscr{A}\zeta(c)$, whence the effect $c \circ \mathscr{A}$ will separate the same states.

Notice that transformations act on effects from the right, inheriting the composition rule of transformations ($\mathscr{B} \circ \mathscr{A}$ means "$\mathscr{A}$ followed by $\mathscr{B}$"). Notice also that $e \circ \mathscr{A} \in [\mathscr{I} \circ \mathscr{A}]_{\text{eff}} = a$. It follows that for $\mathscr{D}$ deterministic one has $\mathscr{D} \in e$, whence $\mathscr{D} \circ \mathscr{A} \in [\mathscr{A}]_{\text{eff}}$.

Consistently, in the "Schrödinger picture," we have $\mathscr{B}\omega(\cdot \circ \mathscr{A}) = \omega(\cdot \circ \mathscr{B} \circ \mathscr{A})$, corresponding to $(\mathscr{B} \circ \mathscr{A})\omega = \omega(\cdot \circ \mathscr{B} \circ \mathscr{A})$. Also, we will use the unambiguous notation $\mathscr{B}\omega(a) = [\mathscr{B}\omega](a)$, whence $\mathscr{B}\omega(a) = \omega(a \circ \mathscr{B})$, and $\omega(a) = \mathscr{A}\omega(e),\ \forall \mathscr{A} \in a$.

5.2.5 Linear structures for transformations and effects

Transformations $\mathscr{A}_1$ and $\mathscr{A}_2$, for which one has the bound $\omega(\mathscr{A}_1) + \omega(\mathscr{A}_2) \leqslant 1$, $\forall \omega \in \mathfrak{S}$, can in principle occur in the same test, and we will call them **test-compatible**. For test-compatible transformations one can define their addition $\mathscr{A}_1 + \mathscr{A}_2$ via the probability rule

$$(\mathscr{A}_1 + \mathscr{A}_2)\omega = \mathscr{A}_1\omega + \mathscr{A}_2\omega, \tag{5.4}$$

where we remind the reader that $\mathscr{A}\omega := \omega(\cdot \circ \mathscr{A})$. Therefore the sum of two test-compatible transformations is just the union-event $\mathscr{A}_1 + \mathscr{A}_2 = \mathscr{A}_1 \cup \mathscr{A}_2$, with the two transformations regarded as belonging to the same test.[18] For any test $\mathbb{A}$ we can define its **total coarse-graining** as the deterministic transformation $\mathscr{D}_{\mathbb{A}} = \sum_{\mathscr{A}_i \in \mathbb{A}} \mathscr{A}_i$. We can trivially extend the addition rule (5.4) to any set of (generally non-test-compatible) transformations, and to subtraction of transformations as well. Notice that the composition "$\circ$" is distributive with respect to addition "$+$."

We can define the multiplication $\lambda\mathscr{A}$ of a transformation $\mathscr{A}$ by a scalar $0 \leqslant \lambda \leqslant 1$ by the rule

$$\omega(\cdot \circ \lambda\mathscr{A}) = \lambda\omega(\cdot \circ \mathscr{A}), \tag{5.5}$$

namely $\lambda\mathscr{A}$ is the transformation conditioning-equivalent to $\mathscr{A}$, but occurring with rescaled probability $\omega(\lambda\mathscr{A}) = \lambda\omega(\mathscr{A})$ – as happens in the convex combination of tests. It follows that for every couple of transformations $\mathscr{A}$ and $\mathscr{B}$ the transformations $\lambda\mathscr{A}$ and $(1 - \lambda)\mathscr{B}$ are test-compatible for $0 \leqslant \lambda \leqslant 1$, consistently with the convex closure of the system S. By extending the definition (5.5) to any positive λ, we then introduce the cone $\mathfrak{T}_+$ of transformations. We will call an event $\mathscr{A}$ **atomic**

[18] The probabilistic class of $\mathscr{A}_1 + \mathscr{A}_2$ is given by

$$\omega(\mathscr{A}_1 + \mathscr{A}_2) = \omega(\mathscr{A}_1) + \omega(\mathscr{A}_2), \quad \forall \omega \in \mathfrak{S},$$

whereas the conditional class is given by

$$\omega_{\mathscr{A}_1 + \mathscr{A}_2} = \frac{\omega(\mathscr{A}_1)}{\omega(\mathscr{A}_1 + \mathscr{A}_2)}\omega_{\mathscr{A}_1} + \frac{\omega(\mathscr{A}_2)}{\omega(\mathscr{A}_1 + \mathscr{A}_2)}\omega_{\mathscr{A}_2}, \quad \forall \omega \in \mathfrak{S}.$$

if it has no non-trivial refinement in any test, namely if it cannot be written as $\mathscr{A} = \sum_i \mathscr{A}_i$ with $\mathscr{A}_i \neq \lambda_i \mathscr{A}$ for some i and $0 < \lambda_i < 1$. Notice that the identity transformation is not necessarily atomic.[19] The set of extremal rays of the cone $\mathfrak{T}_+$ – denoted by $\mathsf{Erays}(\mathfrak{T}_+)$ – contains the atomic transformations.

The notions of (i) test-compatibility, (ii) sum, and (iii) multiplication by a scalar are naturally inherited from transformations to effects via probabilistic equivalence, and then to states via duality. Correspondingly, we introduce the cone of effects $\mathfrak{E}_+$, and, by duality, we extend the cone of states $\mathfrak{S}_+$ to the dual cone of $\mathfrak{E}_+$, completing the set of states $\mathfrak{S}$ to the cone-base of $\mathfrak{S}_+$ made of all positive linear functionals over $\mathfrak{E}_+$ normalized at the deterministic effect, namely all in-principle legitimate states (in parallel we complete the system S with the corresponding state-preparations). We call such a completion of the set of states the **no-restriction hypothesis for states**, corresponding to the **state–effect duality**, namely the convex cones of states $\mathfrak{S}_+$ and of effects $\mathfrak{E}_+$ are dual each other.[20] The state cone $\mathfrak{S}_+$ introduces a natural **partial ordering** $\geqslant$ over states and over effect (via duality), and one has $a \in \mathfrak{E}$ iff $0 \leqslant a \leqslant e$. Thus the convex set $\mathfrak{E}$ **is a truncation of the cone** $\mathfrak{E}_+$, whereas $\mathfrak{S}$ **is a base for the cone** $\mathfrak{S}_+$[21] defined by the normalization condition $\omega \in \mathfrak{S}$ iff $\omega \in \mathfrak{S}_+$ and $\omega(e) = 1$. In the following it will be useful also to express the probability rule $\omega(a)$ also in its dual form $a(\omega) = \omega(a)$, with the effect acting on the state as a linear functional.

By extending (5.5) to any real (complex) scalar λ we build the linear real (complex) span $\mathfrak{T}_{\mathbb{R}} = \mathsf{Span}_{\mathbb{R}}(\mathfrak{T})$ ($\mathfrak{T}_{\mathbb{C}} = \mathsf{Span}_{\mathbb{C}}(\mathfrak{T})$). The *Cartesian decomposition* $\mathfrak{T}_{\mathbb{C}} = \mathfrak{T}_{\mathbb{R}} \oplus \mathrm{i}\mathfrak{T}_{\mathbb{R}}$ holds, i.e., each element $\mathscr{A} \in \mathfrak{T}_{\mathbb{C}}$ can be uniquely written as $\mathscr{A} = \mathscr{A}_R + \mathrm{i}\mathscr{A}_I$, with $\mathscr{A}_R, \mathscr{A}_I \in \mathfrak{T}_{\mathbb{R}}$.[22] Analogously, also for effects and states we define $\mathfrak{E}_{\mathbb{F}}$, $\mathfrak{S}_{\mathbb{F}}$ for $\mathbb{F} = \mathbb{R}, \mathbb{C}$. The state–effect duality implies the linear space identifications $\mathfrak{S}_{\mathbb{F}} \equiv \mathfrak{E}_{\mathbb{F}}$. Thanks to such identifications and to the identity of the dimension of a convex cone with that of its complex and real spans, in the following, without ambiguity, we will simply write $\dim(\mathsf{S}) := \dim[\mathfrak{S}_+(\mathsf{S})] \equiv \dim[\mathfrak{E}_+(\mathsf{S})]$. Moreover, if there is no confusion, then with some abuse of terminology we will simply

[19] For example, the identity transformation is refinable in classical Abelian probabilistic theory, where states are of the form $\varrho = \sum_l p_l |l\rangle\langle l|$, with $\{|l\rangle\}$ a complete orthonormal basis and $\{p_l\}$ a probability distribution. Here the identity transformation is given by $\mathscr{I} = \sum_k |k\rangle\langle k| \cdot |k\rangle\langle k|$, $\{|k\rangle\}$, which is refinable into rank-one projection maps.

[20] In infinite dimensions one also takes the closure of cones.

[21] We remind the reader that a set $\mathsf{B} \subset \mathsf{C}$ of a cone C in a vector space V is called the *base* of C if $0 \notin \mathsf{B}$ and for every point $u \in \mathsf{C}$, $u \neq 0$, there is a unique representation $u = \lambda v$, with $v \in \mathsf{B}$ and $\lambda > 0$. Then, one has that $u \in \mathsf{C}$ spans an extreme ray of C iff $u = \lambda v$, where $\lambda > 0$ and v is an extreme point of B (see Ref. [32]).

[22] Note that the elements $\mathscr{T} \in \mathfrak{T}_{\mathbb{R}}$ can in turn be decomposed *à la* Jordan as $\mathscr{T} = \mathscr{T}_+ - \mathscr{T}_-$, with $\mathscr{T}_\pm \in \mathfrak{T}_+$. However, such a decomposition is generally not unique. According to a theorem of Béllissard and Jochum [33] the Jordan decomposition of the elements of the real span of a cone (with $\mathscr{T}_\pm$ orthogonal in $\mathfrak{T}_{\mathbb{R}}$ Euclidean space) is unique if and only if the cone is self-dual.

refer by "states," "effects," and "transformations" to the respective generalized versions that are elements of the cones, or of their real and complex linear spans.

Note that the cones of states and effects contain the origin, i.e., the null vector of the linear space. For the cone of states one has that $\omega = 0$ iff $\omega(e) = 0$ (since for any effect a one has $0 \leqslant \omega(a) \leqslant \omega(e) = 0$, namely $\omega(a) = 0$). On the other hand, the hyperplane which truncates the cone of effects giving the physical convex set $\mathfrak{E}$ is conveniently characterized using any **internal state** ϑ – i.e., a state that can be written as the convex combination of any state with some other state – by using the following lemma.

Lemma 1. *For any $a \in \mathfrak{E}_+$ one has $a = 0$ iff $\vartheta(a) = 0$ and $a = e$ iff $\vartheta(a) = 1$, with ϑ any internal state.*

Proof. For any state ω one can write $\vartheta = p\omega + (1 - p)\omega'$ with $0 \leqslant p \leqslant 1$ and $\omega' \in \mathfrak{S}$. Then one has $\vartheta(a) = 0$ iff $\omega(a) = 0\ \forall \omega \in \mathfrak{S}$, that is iff $a = 0$. Moreover, one has $\vartheta(a) = 1$ iff $\omega(a) = 1\ \forall \omega \in \mathfrak{S}$, i.e., $a = e$.

5.2.6 Observables and informational completeness

An **observable** $\mathbb{L}$ is a complete set of effects $\mathbb{L} = \{l_i\}$ summing to the deterministic effect as $\sum_{l_i \in \mathbb{L}} l_i = e$, namely l_i are the effects of the events of a test. An observable $\mathbb{L} = \{l_i\}$ is named **informationally complete** for S when each effect can be written as a real linear combination of l_i, namely $\mathsf{Span}_{\mathbb{R}}(\mathbb{L}) = \mathfrak{E}_{\mathbb{R}}(\mathsf{S})$. When the effects of $\mathbb{L}$ are linearly independent the informationally complete observable is named *minimal*. Clearly, since $\mathfrak{E}$ is separating for states, **any informationally complete observable separates states**, that is using an informationally complete observable we can reconstruct also any state $\omega \in \mathfrak{S}(\mathsf{S})$ from the set of probabilities $\omega(l_i)$. The existence of a minimal informationally complete observable constructed from the set of available tests is guaranteed by the following theorem.

Theorem 1. (Existence of minimal informationally complete observable). *It is always possible to construct a minimal informationally complete observable for* S *out of a set of tests of* S.

For the proof see Ref. [17].

In the following we will take a fixed minimal informationally complete observable $\mathbb{L} = \{l_i\}$ as a **reference test**, with respect to which all basis-dependent representations will be defined.

Symmetrically to the notion of an informationally complete observable we have the notion of a **separating set of states** $\mathbb{S} = \{\omega_i\}$, in terms of which one can write any state as a real linear combination of the states $\{\omega_i\}$, namely

$\mathfrak{S}_{\mathbb{R}}(\mathsf{S}) = \mathsf{Span}_{\mathbb{R}}(\mathbb{S})$. Regarded as a test $\mathbb{S} = \{\mathscr{S}_i\} \in \mathsf{S}$ the set of states $\{\omega_i\}$ corresponds to the state-reduction $\mathscr{S}_i\omega = \omega(\mathscr{S}_i)\omega_i$, $\forall\omega \in \mathfrak{S}$. When the corresponding effects $[\mathscr{S}_i]_{\text{eff}}$ form an informationally complete observable the test $\mathbb{S}$ would be an example of the *Quantum Bureau International des Poids et Mesures* of Fuchs [34].

5.2.7 Banach structures

On states $\omega \in \mathfrak{S}$ introduce the **natural norm** $\|\omega\| = \sup_{a\in\mathfrak{E}} \omega(a)$, which extends to the whole linear space $\mathfrak{S}_{\mathbb{R}}$ as $\|\omega\| = \sup_{a\in\mathfrak{E}} |\omega(a)|$. Then, we can introduce the dual norm on effects $\|a\| := \sup_{\omega\in\mathfrak{S}_{\mathbb{R}},\,|\omega\|\leqslant 1} |\omega(a)|$, and then on transformations $\|\mathscr{A}\| := \sup_{b\in\mathfrak{E}_{\mathbb{R}},\|b\|\leqslant 1} \|b \circ \mathscr{A}\|$. Closures in norm (for mathematical convenience) make $\mathfrak{E}_{\mathbb{R}}$ and $\mathfrak{S}_{\mathbb{R}}$ a dual Banach pair, and $\mathfrak{T}_{\mathbb{R}}$ a real Banach algebra.[23] Therefore, all operational quantities can be mathematically represented as elements of such Banach spaces.

5.2.8 The Metric

One can define a **natural distance** between states $\omega, \zeta \in \mathfrak{S}$ as follows:

$$d(\omega, \zeta) := \sup_{l\in\mathfrak{E}} l(\omega) - l(\zeta). \tag{5.6}$$

Lemma 2. *The function (5.6) is a metric on* $\mathfrak{S}$*, and is bounded as* $0 \leqslant d(\omega, \zeta) \leqslant 1$.

Proof. For every effect l, $e - l$ is also a effect, whence

$$\begin{aligned} d(\omega, \zeta) &= \sup_{l\in\mathfrak{E}}(l(\omega) - l(\zeta)) = \sup_{l'\in\mathfrak{E}}((e - l')(\omega) - (e - l')(\zeta)) \\ &= \sup_{l'\in\mathfrak{E}}(l'(\zeta) - l'(\omega)) = d(\zeta, \omega), \end{aligned} \tag{5.7}$$

that is, d is symmetric. On the other hand, $d(\omega, \zeta) = 0$ implies that $\zeta = \omega$, since the two states must give the same probabilities for all transformations. Finally, one has

$$\begin{aligned} d(\omega, \zeta) &= \sup_{l\in\mathfrak{E}}(l(\omega) - l(\theta) + l(\theta) - l(\zeta)) \\ &\leq \sup_{l\in\mathfrak{E}}(l(\omega) - l(\theta)) + \sup_{l\in\mathfrak{E}}(l(\theta) - l(\zeta)) = d(\omega, \theta) + d(\theta, \zeta), \end{aligned} \tag{5.8}$$

[23] An algebra of maps over a Banach space inherits the norm induced by that of the Banach space on which it acts. It is then easy to prove that the closure of the algebra under such a norm is a Banach algebra.

that is, it satisfies the triangular inequality, whence d is a metric. By construction, the distance is bounded as $d(\omega, \zeta) \leqslant 1$, since the maximum value of $d(\omega, \zeta)$ is achieved when $l(\omega) = 1$ and $l(\zeta) = 0$.

The natural distance (5.7) is extended to a metric over $\mathfrak{S}_{\mathbb{R}}$ as $d(\omega, \zeta) = \|\omega - \zeta\|$ with $\|\cdot\|$ the norm over $\mathfrak{S}_{\mathbb{R}}$. Analogously we define the distance between effects as $d(a, b) := \sup_{\omega \in \mathfrak{S}} |\omega(a - b)|$.[24]

A relevant property of the metric in (5.6) is its **monotonicity**, namely that the distance between two states can never increase under deterministic evolution, as established by the following lemma.

Lemma 3. (Monotonicity of the state distance). *For every deterministic physical transformation $\mathscr{D} \in \mathfrak{T}$, one has*

$$d(\mathscr{D}\omega, \mathscr{D}\zeta) \leqslant d(\omega, \zeta). \tag{5.9}$$

Proof. First we notice that since $\mathscr{D} \in \mathfrak{T}$ is a physical transformation, for every effect $a \in \mathfrak{E}$ one has also $a \circ \mathscr{D} \in \mathfrak{E}$, whence $\mathfrak{E} \circ \mathscr{D} \subseteq \mathfrak{E}$. Therefore, we have

$$\begin{aligned} d(\mathscr{D}\omega, \mathscr{D}\zeta) :&= \sup_{a \in \mathfrak{E}} \omega(a \circ \mathscr{D}) - \zeta(a \circ \mathscr{D}) \\ &= \sup_{a \in \mathfrak{E} \circ \mathscr{D}} \omega(a) - \zeta(a) \leqslant \sup_{a \in \mathfrak{E}} \omega(a) - \zeta(a) = d(\omega, \zeta). \end{aligned} \tag{5.10}$$

Notice that we take the transformation deterministic only to assure that $\mathscr{D}\omega$ is itself a state for any ω.

5.2.9 Isometric transformations

A deterministic transformation $\mathscr{U}$ is called *isometric* if it preserves the distance between states, namely

$$d(\mathscr{U}\omega, \mathscr{U}\zeta) \equiv d(\omega, \zeta), \qquad \forall \omega, \zeta \in \mathfrak{S}. \tag{5.11}$$

Lemma 4. *In finite dimensions, all the following properties of a transformation are equivalent: (a) it is isometric for $\mathfrak{S}$; (b) it is isometric for $\mathfrak{E}$; (c) it is an automorphism of $\mathfrak{S}$; and (d) it is an automorphism of $\mathfrak{E}$.*

Proof. By definition a transformation of the convex set (of states or effects) is a linear map of the convex set in itself. A linear isometric map of a set in itself is isometric on the linear span of the set.[25] (Recall that the natural distance between

[24] It is easy to check that such a distance satisfies the trangular inequality.

[25] Interestingly, the Mazur–Ulam theorem states that any surjective isometry (not necessarily linear) between real-normed spaces is affine. Therefore, even if non-linear, it would map convex subsets to convex subsets.

states has been extended to a metric over the whole $\mathfrak{S}_{\mathbb{R}}$.) In finite dimensions an isometry on a normed linear space is diagonalizable [35]. Its eigenvalues must have unit modulus, otherwise it would not be isometric. It follows that it is an orthogonal transformation, and, since it maps the set into itself, it must be a linear automorphism of the set. Therefore, an isometric transformation of a convex set is an automorphism of the convex set.[26]

Now, automorphisms of $\mathfrak{S}$ are isometric for $\mathfrak{E}$, since

$$\begin{aligned} d(a\circ\mathscr{U},b\circ\mathscr{U}) &= \sup_{\omega\in\mathfrak{S}}|\omega((a-b)\circ\mathscr{U})| = \sup_{\omega\in\mathfrak{S}}|(\mathscr{U}\omega)(a-b)| \\ &= \sup_{\omega\in\mathscr{U}\mathfrak{S}}|\omega(a-b)| = \sup_{\omega\in\mathfrak{S}}|\omega(a-b)| = d(a,b), \end{aligned} \tag{5.12}$$

and, similarly, automorphisms of $\mathfrak{E}$ are isometric for $\mathfrak{S}$, since

$$\sup_{a\in\mathfrak{E}}[\omega(a\circ\mathscr{U})-\zeta(a\circ\mathscr{U})] = \sup_{a\in\mathfrak{E}\circ\mathscr{U}}[\omega(a)-\zeta(a)] = d(\omega,\zeta). \tag{5.13}$$

Therefore, automorphisms of $\mathfrak{S}$ are isometric for $\mathfrak{E}$, whence, for the first part of the proof, they are automorphisms of $\mathfrak{E}$, whence they are isometric for $\mathfrak{S}$.

The *physical automorphisms* play the role of unitary transformations in QM.

Corollary 1. (Wigner theorem). *The only transformations of states that are inverted by another transformation must send pure states to pure states, and are isometric.*

5.2.10 The C-algebra of transformations*

We can represent the transformations as elements of $\mathfrak{T}_{\mathbb{C}}$ regarded as a complex C^*-algebra. This is obvious, since $\mathfrak{T}_{\mathbb{C}}$ are by definition linear transformations of effects, making an associative sub-algebra $\mathfrak{T}_{\mathbb{C}}\subseteq\mathsf{Lin}(\mathfrak{E}_{\mathbb{C}})$ of the matrix algebra over $\mathfrak{E}_{\mathbb{C}}$. The **adjoint** and **norm** can be easily defined in terms of any chosen **scalar product** $(\cdot,\cdot)$ **over** $\mathfrak{E}_{\mathbb{C}}$, with the adjoint defined as $(a\circ\mathscr{A}^\dagger,b)=(a,b\circ\mathscr{A})$, and the norm as $\|\mathscr{A}\|=\sup_{a\in\mathfrak{E}_{\mathbb{C}}}\|a\circ\mathscr{A}\|/\|a\|$, with $\|a\|=\sqrt{(a,a)}$. (Notice that these norms are different from the "natural norms" defined in Section 5.2.7.) We can then extend the complex linear space $\mathfrak{T}_{\mathbb{C}}$ by adding the adjoint transformations and taking the norm-closure. We will denote such extension with the same symbol $\mathfrak{T}_{\mathbb{C}}$, which is now a C^*-algebra. Indeed, upon reconstructing $\mathfrak{E}_{\mathbb{C}}$ and $\mathfrak{T}_{\mathbb{C}}$ from the original real spaces via the Cartesian decomposition $\mathfrak{E}_{\mathbb{C}}=\mathfrak{E}_{\mathbb{R}}\oplus\mathrm{i}\mathfrak{E}_{\mathbb{R}}$ and $\mathfrak{T}_{\mathbb{C}}=\mathfrak{T}_{\mathbb{R}}\oplus\mathrm{i}\mathfrak{T}_{\mathbb{R}}$, and introducing the scalar product on $\mathfrak{E}_{\mathbb{C}}$ as the sesquilinear extension of a real symmetric scalar product $(\cdot,\cdot)_{\mathbb{R}}$ over $\mathfrak{E}_{\mathbb{R}}$, the adjoint of a real element $\mathscr{A}\in\mathfrak{T}_{\mathbb{R}}$ is just

[26] For a convex set, an automorphism must send the set to itself keeping the convex structure, whence it must be a one-to-one map that is linear on the span of the convex set.

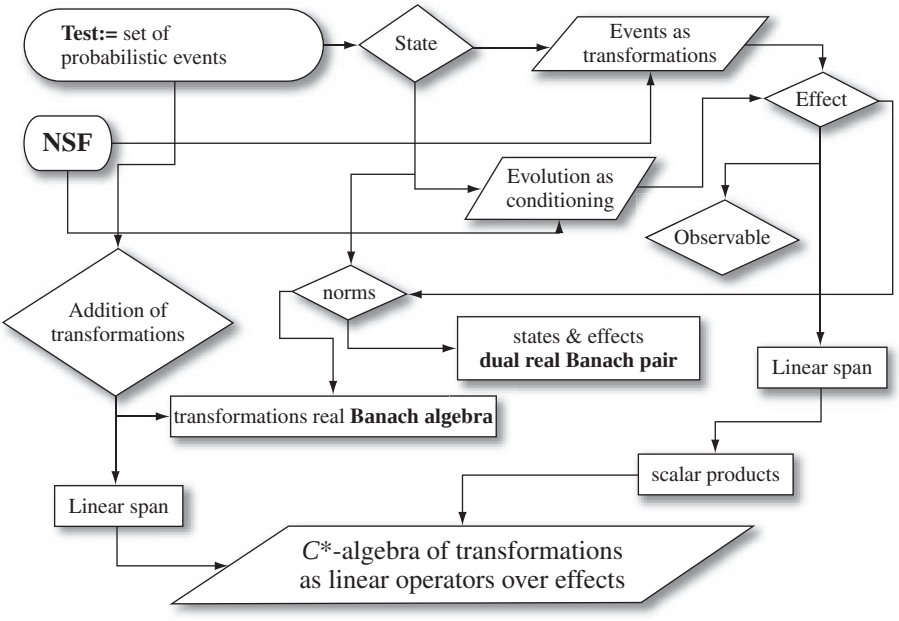


Fig. 5.1 A logical flow chart leading to the representation of any probabilistic theory in terms of a C^*-algebra of linear transformations over the linear space of complex effects (see also footnote 27 and Section 5.3.3 for an operational basis for the scalar product.)

the transposed matrix $\mathscr{A}^t$ with respect to a real basis orthonormal for $(\cdot,\cdot)_{\mathbb{R}}$, and $\mathscr{A}^\dagger := \mathscr{A}_R{}^t - \mathrm{i}\mathscr{A}_I{}^t$ for a general $\mathscr{A} = \mathscr{A}_R + \mathrm{i}\mathscr{A}_I \in \mathfrak{T}_{\mathbb{C}}$. A natural choice of matrix representation for $\mathfrak{T}_{\mathbb{R}}$ is given by its action over a minimal informational complete observable $\mathbb{L} = \{l_i\}$ (the scalar product $(\cdot,\cdot)_{\mathbb{R}} := (\cdot,\cdot)_{\mathbb{L}}$ will correspond to declaring $\mathbb{L}$ as orthonormal). Upon expanding $[l_i \circ \mathscr{A}]_{\text{eff}}$ again over $\mathbb{L} = \{l_i\}$ one has the matrix representation $l_i \circ \mathscr{A} = \sum_j \mathscr{A}_{ji} l_j$. Using the fact that $\mathbb{L}$ is state-separating, we can write the probability rule as the pairing $\omega(a) = (\omega, a)_{\mathbb{R}}$ between $\mathfrak{E}_{\mathbb{R}}$ and $\mathfrak{S}_{\mathbb{R}}$ (and analogously for their complex spans).[27] In this way we see that for every probabilistic theory one can always represent transformations/events as elements of the C^*-algebra $\mathfrak{T}_{\mathbb{C}}$ of matrices acting on the linear space of complex effects $\mathfrak{E}_{\mathbb{C}}$. In Figure 5.1 the logical derivation of the C^*-algebra representation of the theory is summarized.

[27] The present derivation of the C^*-algebra representation of transformations is more direct than that in Ref. [17], and is just equivalent to the probabilistic framework inherent in the notion of a "test" (see also the summary of the whole logical deduction in the flow chart in Figure 5.1). The specific C^*-algebra in Ref. [17] possessed operational notions of adjoint and of scalar product over effects, both constructed using a symmetric faithful bipartite state, needing in this way two additional postulates: (a) the existence of dynamically independent systems and (b) the existence of faithful symmetric bipartite states. Such construction is briefly reviewed in Section 5.3.3.

Conversely, given (1) a C^*-algebra $\mathfrak{T}_{\mathbb{C}}$, (2) the cone of transformations $\mathfrak{T}_+$, and (3) the vector $e \in \mathfrak{E}_{\mathbb{C}}$ representing the deterministic effect, we can rebuild the full probabilistic theory by constructing the cone of effects as the orbit $\mathfrak{E}_+ = e \circ \mathfrak{T}_+$, and taking the cone of states $\mathfrak{S}_+$ as the dual cone of $\mathfrak{E}_+$.[28]

5.3 Independent systems

5.3.1 Dynamical independence and marginal states

A purely dynamical notion of *system independence* coincides with the possibility of performing local tests. To be precise, we will call systems S_1 and S_2 **independent** if it is possible to perform their tests as **local tests**, i.e., in such a way that for every joint state of S_1 and S_2 the transformations on S_1 commute with transformations on S_2, namely[29]

$$\mathscr{A}^{(1)} \circ \mathscr{B}^{(2)} = \mathscr{B}^{(2)} \circ \mathscr{A}^{(1)}, \ \forall \mathscr{A}^{(1)} \in \mathbb{A}^{(1)}, \ \forall \mathscr{B}^{(2)} \in \mathbb{B}^{(2)}. \tag{5.14}$$

The local tests comprise the Cartesian product $\mathsf{S}_1 \times \mathsf{S}_2$, which is closed under cascade. We will close this set also under convex combination, coarse-graining, and conditioning, making it a "system," denote such a system with the same symbol $\mathsf{S}_1 \times \mathsf{S}_2$, and call **local** all tests in $\mathsf{S}_1 \times \mathsf{S}_2$. We now **compose** the two systems S_1 and S_2 into the **bipartite** system $\mathsf{S}_1 \odot \mathsf{S}_2$ by adding the local tests into the new system $\mathsf{S}_1 \odot \mathsf{S}_2$ as $\mathsf{S}_1 \odot \mathsf{S}_2 \supseteq \mathsf{S}_1 \times \mathsf{S}_2$ and closing under cascading, coarse-graining, and convex combination. We call the tests in $\mathsf{S}_1 \odot \mathsf{S}_2 \backslash \mathsf{S}_1 \times \mathsf{S}_2$ **non-local**, and we will extend the local/non-local nomenclature to the pertaining transformations. In the following for identical systems we will also use the notation $\mathsf{S}^{\odot N} = \mathsf{S} \odot \mathsf{S} \odot \ldots \odot \mathsf{S}$ (N times), and $\mathfrak{Z}^{\odot N} := \mathfrak{Z}(\mathsf{S}^{\odot N})$ to denote N-partite sets/spaces, with $\mathfrak{Z} = \mathfrak{S}, \mathfrak{S}_+, \mathfrak{S}_{\mathbb{R}}, \mathfrak{S}_{\mathbb{C}}, \mathfrak{E}, \mathfrak{E}_+, \ldots$.

Since the local transformations commute, we will just put them in a string, as $(\mathscr{A}, \mathscr{B}, \mathscr{C}, \ldots) := \mathscr{A}^{(1)} \circ \mathscr{A}^{(2)} \circ \mathscr{A}^{(3)} \circ \ldots$ (convex combinations and coarse graining will be sums of strings). Clearly, since the probability $\omega(\mathscr{A}, \mathscr{B}, \mathscr{C}, \ldots)$ is independent of the time ordering of transformations, it is just a function only of the effects $\omega(\mathscr{A}, \mathscr{B}, \mathscr{C}, \ldots) = \omega([\mathscr{A}]_{\text{eff}}, [\mathscr{B}]_{\text{eff}}, [\mathscr{C}]_{\text{eff}}, \ldots)$, namely the joint effect corresponding to local transformations is made of (sums of) local effects $[(\mathscr{A}, \mathscr{B}, \mathscr{C}, \ldots)]_{\text{eff}} \equiv ([\mathscr{A}]_{\text{eff}}, [\mathscr{B}]_{\text{eff}}, [\mathscr{C}]_{\text{eff}}, \ldots)$.

The embedding of local tests $\mathsf{S}_1 \times \mathsf{S}_2$ into the bipartite system $\mathsf{S}_1 \odot \mathsf{S}_2$ implies that $\mathfrak{T}_{\mathbb{F}}(\mathsf{S}_1 \odot \mathsf{S}_2) \supseteq \mathfrak{T}_{\mathbb{F}}(\mathsf{S}_1) \otimes \mathfrak{T}_{\mathbb{F}}(\mathsf{S}_2)$ and $\mathfrak{E}_{\mathbb{F}}(\mathsf{S}_1 \odot \mathsf{S}_2) \supseteq \mathfrak{E}_{\mathbb{F}}(\mathsf{S}_1) \otimes \mathfrak{E}_{\mathbb{F}}(\mathsf{S}_2)$,

[28] The "orbit" $e \circ \mathfrak{T}_+$ is defined as the set $e \circ \mathfrak{T}_+ := \{e \circ \mathscr{A} | \mathscr{A} \in \mathfrak{T}_+\}$.

[29] The present definition of independent systems is purely dynamical, in the sense that it does not involve statistical requirements, e.g., the existence of factorized states. This, however, is implied by the mentioned no-restriction hypothesis for states.

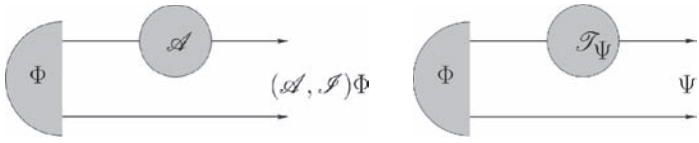


Fig. 5.2 Illustrations of the notions of a dynamically (left) and a preparationally (right) faithful state for a bipartite system. A bipartite state Φ is dynamically faithful with respect to system S_1 when the output state $(\mathscr{A}, \mathscr{I})\Phi$ is in one-to-one correspondence with the local transformation $\mathscr{A}$ on system S_1, whereas it is preparationally faithful with respect to S_1 if every bipartite state Ψ can be achieved as $\Psi = (\mathscr{T}_\Psi, \mathscr{I})\Phi$ via a local transformation $\mathscr{T}_\Psi$ on S_1.

both for real and for complex spans $\mathbb{F} = \mathbb{R}, \mathbb{C}$. On the other hand, since local tests include local state-preparation (or, otherwise, because of the no-restriction hypothesis for states) the set of bipartite states $\mathfrak{S}(\mathsf{S}_1 \odot \mathsf{S}_2)$ always includes the **factorized states**, i.e., those corresponding to factorized probability rules, e.g., $\Omega(a, b) = \omega_1(a)\omega_2(b)$ for local effects a and b. In parallel with local transformations and effects, we will denote factorized states as strings $\Omega = (\omega_1, \omega_2, \ldots)$, e.g., $(\omega_1, \omega_2)(a, b) = \omega_1(a)\omega_2(b)$. Then, closure under convex combination implies that $\mathfrak{S}_\mathbb{F}(\mathsf{S}_1 \odot \mathsf{S}_2) \supseteq \mathfrak{S}_\mathbb{F}(\mathsf{S}_1) \otimes \mathfrak{S}_\mathbb{F}(\mathsf{S}_2)$, for $\mathbb{F} = \mathbb{R}, \mathbb{C}$.

For N systems in the joint state Ω, we define the **marginal state** $\Omega|_n$ of the nth system as the probability rule for any local transformation $\mathscr{A}$ at the nth system, with all other systems untouched, namely

$$\Omega|_n(\mathscr{A}) \doteq \Omega(\mathscr{I}, \ldots, \mathscr{I}, \underbrace{\mathscr{A}}_{n\text{th}}, \mathscr{I}, \ldots). \tag{5.15}$$

Clearly, since the probability for local transformations depends only on their respective effects, the marginal state is equivalently defined as

$$\Omega|_n(a) \doteq \Omega(e, \ldots, e, \underbrace{a}_{n\text{th}}, e, \ldots) \quad \text{for } a \in \mathfrak{E}. \tag{5.16}$$

It readily follows that the marginal state $\Omega|_n$ is independent of any deterministic transformation – i.e., any test – that is performed on systems different from the nth: this is exactly the general statement of the **no-signaling** condition or **acausality of local tests**. Therefore, the present notion of dynamical independence directly implies the no-signaling condition. The definition in (5.15) can be trivially extended to unnormalized states.[30,31]

[30] Notice that any generally unnormalized state is zero iff the joint state is zero, since $\Omega(e, e, \ldots, e) = \Omega_n(e) = 0$.

[31] The present notion of dynamical independence is indeed so minimal that it can be satisfied not only by the quantum tensor product, but also by the quantum direct sum [36]. (Notice, however, that an analogue of Tsirelson's theorem [37] for transformations in finite dimensions would imply a representation of dynamical independence over the tensor product of effects.) In order to extract only the tensor product an additional assumption is needed. As shown in Refs. [17, 36] two possibilities are either postulating the existence of

In the following we will use the following identities:

$$\Psi|_2(a) = \Psi(e, a) = \Psi(e, e \circ \mathscr{A}) = (\mathscr{I}, \mathscr{A})\Psi(e, e), \ \forall \mathscr{A} \in a. \tag{5.17}$$

5.3.2 Faithful states

A bipartite state $\Phi \in \mathfrak{S}(\mathsf{S}_1 \odot \mathsf{S}_2)$ is **dynamically faithful** with respect to S_1 when the output state $(\mathscr{A}, \mathscr{I})\Phi$ is in one-to-one correspondence with the local transformation $\mathscr{A}$ on system S_1, that is, the cone-homomorphism[32] $\mathscr{A} \leftrightarrow (\mathscr{A}, \mathscr{I})\Phi$ from $\mathfrak{T}_+(\mathsf{S}_1)$ to $\mathfrak{S}_+(\mathsf{S}_1 \odot \mathsf{S}_2)$ is a monomorphism.[33] Equivalently the map $\mathscr{A} \mapsto (\mathscr{A}, \mathscr{I})\Phi$ extends to an injective linear map between the linear spaces $\mathfrak{T}_{\mathbb{R}}(\mathsf{S}_1)$ and $\mathfrak{S}_{\mathbb{R}}(\mathsf{S}_1 \odot \mathsf{S}_2)$ preserving the partial ordering relative to the spanning cones, and this is true also in the inverse direction on the range of the map. Notice that no physical transformation $\mathscr{A} \neq 0$ "annihilates" Φ, i.e., gives $(\mathscr{A}, \mathscr{I})\Phi = 0$.

A bipartite state $\Phi \in \mathfrak{S}(\mathsf{S}_1 \odot \mathsf{S}_2)$ is called **preparationally faithful** with respect to S_1 if every bipartite state Ψ can be achieved as $\Psi = (\mathscr{T}_\Psi, \mathscr{I})\Phi$ by a local transformation $\mathscr{T}_\Psi \in \mathfrak{T}_+(\mathsf{S}_1)$. This means that the cone-homomorphism $\mathscr{A} \mapsto (\mathscr{A}, \mathscr{I})\Phi$ from $\mathfrak{T}_+(\mathsf{S}_1)$ to $\mathfrak{S}_+(\mathsf{S}_1 \odot \mathsf{S}_2)$ is an epimorphism. Equivalently, the map $\mathscr{A} \mapsto (\mathscr{A}, \mathscr{I})\Phi$ extends to a surjective linear map between the linear spaces $\mathfrak{T}_{\mathbb{R}}(\mathsf{S}_1)$ and $\mathfrak{S}_{\mathbb{R}}(\mathsf{S}_1 \odot \mathsf{S}_2)$ preserving the partial ordering relative to the spanning cones.

In simple words, a dynamically faithful state keeps the imprinting of a local transformation on the output, i.e., from the output we can recover the transformation. On the other hand, a preparationally faithful state allows us to prepare any desired joint state (probabilistically) by means of local transformations. Dynamical and preparational faithfulness correspond to the properties of being *separating* and *cyclic* for the C^*-algebra of transformations.

Theorem 2. *The following assertions hold.*

(1) *Any state* $\Phi \in \mathfrak{S}(\mathsf{S}_1 \odot \mathsf{S}_2)$ *that is preparationally faithful with respect to* S_1 *is dynamically faithful with respect to* S_2.

(2) *For identical systems in finite dimensions any state* Φ *that is preparationally faithful with respect to a system is also dynamically faithful with respect to the same system,*

bipartite states that are dynamically and preparationally faithful, or postulating the local observability principle. Here we will consider the former as a postulate, and derive the latter as a theorem.

[32] A cone-homomorphism between cones C_1 and C_2 is a linear map between $\mathsf{Span}_{\mathbb{R}}(\mathsf{C}_1)$ and $\mathsf{Span}_{\mathbb{R}}(\mathsf{C}_2)$ that sends elements of C_1 to elements of C_2, but not necessarily vice versa.

[33] This means that $(\mathscr{A}_1, \mathscr{I})\Phi = (\mathscr{A}_2, \mathscr{I})\Phi$ iff $\mathscr{A}_1 = \mathscr{A}_2$, or, in other words, $\forall \mathscr{A} \in \mathfrak{T}_{\mathbb{R}}$: $(\mathscr{A}, \mathscr{I})\Phi = 0 \iff \mathscr{A} = 0$.

and one has the cone-isomorphism[34] $\mathfrak{T}_+(\mathsf{S}) \simeq \mathfrak{S}_+(\mathsf{S}^{\odot 2})$. *Moreover, a local transformation on* Φ *produces an output pure (unnormalized) bipartite state iff the transformation is atomic.*

(3) *If there exists a state of* $\mathsf{S}_1 \odot \mathsf{S}_2$ *that is preparationally faithful with respect to* S_1, *then* $\dim(\mathsf{S}_1) \geqslant \dim(\mathsf{S}_2)$.

(4) *If there exists a state of* $\mathsf{S}_1 \odot \mathsf{S}_2$ *that is preparationally faithful with respect to both systems, then one has the cone-isomorphisms* $\mathfrak{E}_+(\mathsf{S}_1) \simeq \mathfrak{S}_+(\mathsf{S}_2)$ *and* $\mathfrak{E}_+(\mathsf{S}_2) \simeq \mathfrak{S}_+(\mathsf{S}_1)$.

(5) *If for two identical systems there exists a state that is preparationally faithful with respect to both systems, then one has the cone-isomorphism* $\mathfrak{S}_+ \simeq \mathfrak{E}_+$ *(weak self-duality).*

(6) *If the state* $\Phi \in \mathfrak{S}(\mathsf{S}_1 \odot \mathsf{S}_2)$ *is preparationally faithful with respect to* S_1, *then for any invertible transformation* $\mathscr{A} \in \mathfrak{T}_+(\mathsf{S}_1)$ *also the (unnormalized) state* $(\mathscr{A}, \mathscr{I})\Phi$ *is preparationally faithful with respect to the same system. In particular, it will be a faithful state for any physical automorphism of* $\mathfrak{S}(\mathsf{S}_1)$.[35]

(7) *For identical systems in finite dimensions, for* Φ *preparationally faithful with respect to both systems, the state* $\chi := \Phi(e, \cdot)$ *is cyclic in* $\mathfrak{S}_+(\mathsf{S})$ *under* $\mathfrak{T}_+(\mathsf{S})$, *and the observables* $\mathbb{L} = \{l_i\}$ *of* S_2 *are in one-to-one correspondence with the ensemble decompositions* $\{\rho_i\}_{i=1}^{|\mathbb{L}|}$ *of* χ, *with* $\rho_i := \Phi(l_i, \cdot)$, *and* χ *is an internal state.*

Proof.

(1) Introduce the map $\omega \mapsto \mathscr{T}_\omega$ where for every $\omega \in \mathfrak{S}(\mathsf{S}_2)$ one chooses a local transformation $\mathscr{T}_\omega$ on S_1 such that $(\mathscr{T}_\omega, \mathscr{I})\Phi|_2 = \omega$. This is possible because Φ is preparationally faithful with respect to S_1. One has $\mathscr{A}\omega = (\mathscr{T}_\omega, \mathscr{A})\Phi|_2 = (\mathscr{T}_\omega, \mathscr{I})(\mathscr{I}, \mathscr{A})\Phi|_2 \ \forall \omega \in \mathfrak{S}(\mathsf{S}_2)$. Therefore, from $(\mathscr{I}, \mathscr{A})\Phi$ one can recover the action of $\mathscr{A}$ on any state ω by first applying $(\mathscr{T}_\omega, \mathscr{I})$ and then take the marginal, i.e., one recovers $\mathscr{A}$ from $(\mathscr{I}, \mathscr{A})\Phi$, which is another way of saying that $\mathscr{A} \mapsto (\mathscr{I}, \mathscr{A})\Phi$ is injective, namely Φ is dynamically faithful with respect to S_2.

(2) Denote by $\Phi \in \mathfrak{S}^{\odot 2}$ a state that is preparationally faithful with respect to S_1. Since the linear map $\mathscr{A} \mapsto (\mathscr{A}, \mathscr{I})\Phi$ from $\mathfrak{T}_\mathbb{R}$ to $\mathfrak{S}_\mathbb{R}^{\odot 2}$ is surjective, one has $\dim(\mathfrak{T}_\mathbb{R}) \geqslant \dim(\mathfrak{S}_\mathbb{R}^{\odot 2})$. However, one has also $\dim(\mathfrak{T}_\mathbb{R}) \leqslant \dim(\mathfrak{S}_\mathbb{R}^{\odot 2})$ since $\mathfrak{T}_\mathbb{R} \subseteq \mathsf{Lin}(\mathfrak{S}_\mathbb{R}) \simeq \mathfrak{S}_\mathbb{R}^{\otimes 2} \subseteq \mathfrak{S}_\mathbb{R}^{\odot 2}$, whence $\dim(\mathfrak{T}_\mathbb{R}) = \dim(\mathfrak{S}_\mathbb{R}^{\odot 2})$, and, having null kernel, the map is also injective, whence Φ is dynamically faithful with respect to S_1. Since now the state Φ is both preparationally and dynamically faithful with respect to the same system S_1, it

[34] We say that two cones C_1 and C_2 are isomorphic (denoted as $\mathsf{C}_1 \simeq \mathsf{C}_2$) if there exists a one-to-one linear mapping between $\mathsf{Span}_\mathbb{R}(\mathsf{C}_1)$ and $\mathsf{Span}_\mathbb{R}(\mathsf{C}_2)$ that is cone-preserving in both directions. We will call such a map a cone-isomorphism between the two cones. Such a map will send extremal rays of C_1 to extremal rays of C_2 and positive linear combinations to positive linear combinations, and the same is true for the inverse map.

[35] One may be tempted to consider all automorphisms of $\mathfrak{S}(\mathsf{S}_1)$, instead of just the physical ones. However, there is no guarantee that any automorphism will be also an automorphism of bipartite states when applied locally. This is the case of QM, where the transposition is an automorphism of $\mathfrak{S}(\mathsf{S}_1)$, and nevertheless is not a local automorphism of $\mathfrak{S}(\mathsf{S}_1 \odot \mathsf{S}_2)$.

follows that the map $\mathscr{A} \mapsto (\mathscr{A}, \mathscr{I})\Phi$ establishes the cone-isomorphism $\mathfrak{T}_+ \simeq \mathfrak{S}_+^{\odot 2}$. Since the faithful state establishes the cone-isomorphism $\mathfrak{T}_+ \simeq \mathfrak{S}_+^{\odot 2}$, it maps extremal rays of $\mathfrak{T}_+$ to extremal rays of $\mathfrak{S}_+^{\odot 2}$ and vice versa; that is, $\mathscr{A} \in \mathsf{Erays}(\mathfrak{T}_+)$ iff $(\mathscr{A}, \mathscr{I})\Phi \in \mathsf{Erays}(\mathfrak{S}_+^{\odot 2})$.

(3) For Φ preparationally faithful with respect to S_1, consider the cone homomorphism $a \mapsto \omega_a := \Phi(a, \cdot)$ which associates an (unnormalized) state $\omega_a \in \mathfrak{S}_+(\mathsf{S}_2)$ with each effect $a \in \mathfrak{E}_+(\mathsf{S}_1)$. The extension to a linear map $a \mapsto \omega_a$ between the linear spaces $\mathfrak{S}_{\mathbb{R}}(\mathsf{S}_2)$ and $\mathfrak{E}_{\mathbb{R}}(\mathsf{S}_1)$ preserves the cone structure, and is surjective, since Φ is preparationally faithful with respect to S_1 (whence every bipartite state, and, in particular, every marginal state, can be obtained from a local effect). The bound $\dim(\mathsf{S}_1) \geqslant \dim(\mathsf{S}_2)$ then follows from surjectivity.

(4) Similarly to the proof of item (1), consider the map $\lambda \mapsto \mathscr{T}_\lambda$, where for every marginal state $\lambda \in \mathfrak{S}(\mathsf{S}_1)$ one chooses a local transformation $\mathscr{T}_\lambda$ on S_2 such that $(\mathscr{I}, \mathscr{T}_\lambda)\Phi|_1 = \lambda$ (Φ is preparationally faithful with respect to S_2). Then, one has

$$\forall \lambda \in \mathfrak{S}(\mathsf{S}_1),\ \lambda(a) = (\mathscr{I}, \mathscr{T}_\lambda)\Phi(a, e) = \Phi(a, \mathscr{T}_\lambda) = \omega_a(\mathscr{T}_\lambda). \tag{5.18}$$

It follows that $\omega_a = \omega_b$ implies that $\lambda(a) = \lambda(b)$ for all states $\lambda \in \mathfrak{S}(\mathsf{S}_1)$; that is, $a = b$, whence the homomorphism $a \mapsto \omega_a$ which is surjective (since Φ is preparationally faithful) is also injective, i.e., is bijective, and, since it maps elements of $\mathfrak{E}_+(\mathsf{S}_1)$ to elements of $\mathfrak{S}_+(\mathsf{S}_2)$ and, vice versa, to each element of $\mathfrak{S}_+(\mathsf{S}_2)$, it corresponds to an element of $\mathfrak{E}_+(\mathsf{S}_1)$ (Φ is preparationally faithful), thus it is a cone-isomorphism. We then have the cone-isomorphism $\mathfrak{E}_+(\mathsf{S}_1) \simeq \mathfrak{S}_+(\mathsf{S}_2)$. The cone-isomorphism $\mathfrak{E}_+(\mathsf{S}_2) \simeq \mathfrak{S}_+(\mathsf{S}_1)$ follows on exchanging the two systems.

(5) According to point (4) one has the cone-isomorphism $\mathfrak{E}_+(\mathsf{S}_1) \simeq \mathfrak{S}_+(\mathsf{S}_2) \simeq \mathfrak{S}_+(\mathsf{S}_1)$.

(6) This is obvious, from the definition of a preparationally faithful state.

(7) According to (4) $\omega_a := \Phi(a, \cdot)$ establishes the cone-isomorphism $\mathfrak{E}_+(\mathsf{S}) \simeq \mathfrak{S}_+(\mathsf{S})$. On the other hand, since the state is both preparationally and dynamically faithful for either system, then for any transformation $\mathscr{T}$ on the first system there exists a unique transformation $\mathscr{T}'$ on the other system giving the same output state (see also the definition of the "transposed" transformation with respect to a dynamically faithful state in the following). Therefore, since any effect a can be written as $a = e \circ \mathscr{T}_a$ for any $\mathscr{T}_a \in a$, one has $\omega_a = \Phi(e \circ \mathscr{T}_a, \cdot) = \Phi(e, \cdot \circ \mathscr{T}'_a) = \mathscr{T}'_a \chi$. The observable–ensemble correspondence and the fact that χ is an internal state are both immediate consequences of the fact that $\omega_a := \Phi(a, \cdot)$ is a cone-isomorphism.

The transposed of a transformation (Figure 5.3). For a symmetric bipartite state Φ of two identical systems that is preparationally faithful for one system – hence, according to Theorem 2, is both dynamically and preparationally faithful with respect to both systems – one can define operationally the **transposed** $\mathscr{T}'$ of a transformation $\mathscr{T} \in \mathfrak{T}_{\mathbb{R}}$ through the identity

$$\Phi(a, b \circ \mathscr{T}) = \Phi(a \circ \mathscr{T}', b), \tag{5.19}$$

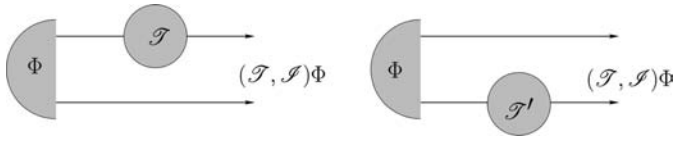


Fig. 5.3 An illustration of the notion of the transposed of a transformation for a symmetric dynamically and preparationally faithful state.

i.e., $(\mathscr{T}', \mathscr{I})\Phi = (\mathscr{I}, \mathscr{T})\Phi$, namely, operationally the transposed $\mathscr{T}'$ of a transformation $\mathscr{T}$ is the transformation which will give the same output bipartite state of $\mathscr{T}$ if operated on the twin system. It is easy to verify (using the symmetry of Φ) that $\mathscr{T}'' = \mathscr{T}$ and that $(\mathscr{B} \circ \mathscr{A})' = \mathscr{A}' \circ \mathscr{B}'$.

We are now in position to formulate the main postulate.

Postulate PFAITH (Existence of a symmetric preparationally faithful pure state). *For any couple of identical systems, there exists a symmetric (under permutation of the two systems) pure state that is preparationally faithful.*

Theorem 2 guarantees that such a state is both dynamically and preparationally faithful, and with respect to both systems, as a consequence of symmetry.[36] Postulate PFAITH thus guarantees that to any system we can adjoin an ancilla and prepare a pure state that is dynamically and preparationally faithful with respect to our system. This is operationally crucial in guaranteeing the preparability of any quantum state for any bipartite system using only local transformations, and to assure the possibility of experimental calibrability of tests for any system. Notice that it would be impossible, even in principle, to calibrate transformations without a dynamically faithful state, since any set of input states $\{\omega_n\} \in \mathsf{S}'$ that is "separating" for transformations $\mathfrak{T}(\mathsf{S}')$ is equivalent to a bipartite state $\Phi = \sum_n \omega_n \otimes \lambda_n \in \mathfrak{S}(\mathsf{S}' \odot \mathsf{S}'')$ that is dynamically faithful for S', with the states $\{\lambda_n\}$ working just as "flags" representing the "knowledge" of which state of the set $\{\omega_n\}$ has been prepared. Notice that in QM every maximal Schmidt-number entangled state of two identical systems is both preparationally and dynamically faithful for both systems. In classical mechanics, on the other hand, a state of the form $\Phi = \sum_l |l\rangle\langle l| \otimes |l\rangle\langle l|$ with $\{|l\rangle\}$ a complete orthogonal set of states (see footnote 19) will be both dynamically and preparationally faithful; however, being not pure, it would require a (possibly unlimited) sequence of preparations.

On the mathematical side, instead, according to Theorem 2 Postulate PFAITH restricts the theory to the weakly self-dual scenario (i.e., with the

[36] In fact, upon denoting by $\mathscr{T}_\Psi$ the local transformation such that $(\mathscr{T}, \mathscr{I})\Phi = \Psi$, one has $(\mathscr{I}, \mathscr{T}_{\mathscr{S}\Psi})\Phi = \Psi$, $\mathscr{S}$ denoting the transformation swapping the two systems.

cone-isomorphism $\mathfrak{S}_+ \simeq \mathfrak{E}_+$), and in finite dimensions one also has the cone-isomorphism $\mathfrak{T}_+(\mathsf{S}) \simeq \mathfrak{S}_+(\mathsf{S}^{\odot 2})$. In addition, one also has the following very useful lemma.

Lemma 5. *For finite dimensions Postulate PFAITH implies that the linear space of transformations is full, i.e.,* $\mathfrak{T}_{\mathbb{F}} = \mathsf{Lin}(\mathfrak{E}_{\mathbb{F}})$. *Moreover, one has* $\mathfrak{S}_{\mathbb{F}}(\mathsf{S}^{\odot 2}) = \mathfrak{S}_{\mathbb{F}}(\mathsf{S})^{\otimes 2}$ *and* $\mathfrak{E}_{\mathbb{F}}(\mathsf{S}^{\odot 2}) = \mathfrak{E}_{\mathbb{F}}(\mathsf{S})^{\otimes 2}$ *for* $\mathbb{F} = \mathbb{R}, \mathbb{C}$, *that is, bipartite states and effects are cones spanning the tensor products* $\mathfrak{S}_{\mathbb{F}}^{\otimes 2}$ *and* $\mathfrak{E}_{\mathbb{F}}^{\otimes 2}$, *respectively.*

Proof. In the following we restrict to finite dimensions, with $\mathbb{F} = \mathbb{R}, \mathbb{C}$ denoting either the real or the complex fields, respectively. According to item (2) of Theorem 2, for two identical systems the existence of a state that is preparationally faithful with respect to either one of the two systems implies $\mathfrak{S}_{\mathbb{F}}(\mathsf{S}^{\odot 2}) \simeq \mathfrak{T}_{\mathbb{F}}(\mathsf{S})$. Since transformations act linearly over effects, one has $\mathfrak{T}_{\mathbb{F}} \subseteq \mathsf{Lin}(\mathfrak{E}_{\mathbb{F}}) \simeq \mathfrak{E}_{\mathbb{F}}^{\otimes 2}$, whence $\mathfrak{E}_{\mathbb{F}}(\mathsf{S}^{\odot 2}) \simeq \mathfrak{S}_{\mathbb{F}}(\mathsf{S}^{\odot 2}) \simeq \mathfrak{T}_{\mathbb{F}}(\mathsf{S}) \subseteq \mathfrak{E}_{\mathbb{F}}(\mathsf{S})^{\otimes 2}$. However, by local-test embedding one also has $\mathfrak{E}_{\mathbb{F}}(\mathsf{S}^{\odot 2}) \supseteq \mathfrak{E}_{\mathbb{F}}(\mathsf{S})^{\otimes 2}$, whence $\mathfrak{E}_{\mathbb{F}}(\mathsf{S}^{\odot 2}) = \mathfrak{E}_{\mathbb{F}}(\mathsf{S})^{\otimes 2}$, which implies that $\mathfrak{T}_{\mathbb{F}} = \mathsf{Lin}(\mathfrak{E}_{\mathbb{F}})$. Finally, by virtue of state-effect duality one also has $\mathfrak{S}_{\mathbb{F}}(\mathsf{S}^{\odot 2}) = \mathfrak{S}_{\mathbb{F}}^{\otimes 2}(\mathsf{S})$.

The above lemma could have been extended to couples of different systems. However, this would necessitate the consideration of more general transformations between different systems (see footnote 13).

We conclude that Postulate PFAITH – i.e., the existence of a symmetric preparationally faithful pure state for bipartite systems – guarantees that we can represent bipartite quantities (states, effects, transformations) as elements of the tensor product of the single-system spaces. This fact also implies the following relevant principle.

Corollary 2. (Local observability principle). *For every composite system there exist informationally complete observables made of local informationally complete observables.*

Proof. A joint observable made of local observables $\mathbb{L} = \{l_i\}$ on S_1 and $\mathbb{M} = \{m_j\}$ on S_2 is of the form $\mathbb{L} \times \mathbb{M} = \{(l_i, m_j)\}$. Then, by definition, the statement of the corollary is $\mathfrak{E}_{\mathbb{R}}(\mathsf{S}^{\odot 2}) \subseteq \mathsf{Span}_{\mathbb{R}}(\mathbb{L} \times \mathbb{M}) = \mathfrak{E}_{\mathbb{R}}^{\otimes 2}(\mathsf{S})$, which is true according to Lemma 5.

Operationally, the local-observability principle plays a crucial role, since it reduces enormously experimental complexity, by guaranteeing that only local (although jointly executed) tests are sufficient to retrieve complete information on a composite system, including all correlations between the components. This principle reconciles holism with reductionism in a non-local theory, in the sense that we can observe a holistic nature in a reductionistic way, i.e., locally.

In addition to Lemma 5 and to the local-observability principle, Postulate PFAITH has a long list of remarkable consequences for the probabilistic theory, which are given by the following theorem.

Theorem 3. *If PFAITH holds, the following assertions are true.*

(1) *The identity transformation is atomic.*
(2) *One has* $\omega_{a\circ\mathscr{A}'} = \mathscr{A}\omega_a$, *or equivalently* $\mathscr{A}\omega = \Phi(a_\omega \circ \mathscr{A}', \cdot)$, *where* $\mathscr{A}'$ *denotes the transposed of* $\mathscr{A}$ *with respect to* Φ.
(3) *The transposed of a physical automorphism of the set of states is still a physical automorphism of the set of states.*
(4) *The marginal state* χ *is invariant under the transposed of a channel (deterministic transformation) and hence, in particular, under a physical automorphism of the set of states.*
(5) *Alice can perform perfect EPR-cheating in a perfect concealing bit-commitment protocol.*

Proof.

(1) According to Theorem 2 item (2), the map $\mathscr{A} \mapsto (\mathscr{A}, \mathscr{I})\Phi$ establishes the cone-isomorphism $\mathfrak{T}_+ \simeq \mathfrak{S}_+^{\odot 2}$, whence on mapping extremal rays of $\mathfrak{T}_+$ to extremal rays of $\mathfrak{S}_+^{\odot 2}$ and vice versa it maps the state Φ itself (which is pure) to the identity, which then must be atomic.
(2) Immediate definition of the transposition with respect to the dynamically faithful state Φ.
(3) Point (2) establishes that the transposed of a state-automorphism is an effect automorphism, which, due to the cone-isomorphism, is again a state-automorphism (see also footnote 35).
(4) For deterministic $\mathscr{T}$ one has $\mathscr{T}'\chi = \Phi(e, \cdot \circ \mathscr{T}') = \Phi(e \cdot \mathscr{T}, \cdot) = \Phi(e, \cdot) = \chi$. The last statement follows from (3) (see also footnote 35).
(5) (For the definition of the protocol, see Ref. [38]). For the protocol to be concealing there must exist two ensembles of states $\{\rho_i^{\mathbb{A}}\}$ and $\{\rho_i^{\mathbb{B}}\}$ that are indistinguishable by Bob. For $\sum_i \rho_i^{\mathbb{A}} = \sum_i \rho_i^{\mathbb{B}} = \chi$ these correspond to the two observables $\mathbb{A} = \{a_i\}$ and $\mathbb{B} = \{b_i\}$ with $\rho_i^{\mathbb{A}} = \Phi(a_i, \cdot)$ and $\rho_i^{\mathbb{B}} = \Phi(b_i, \cdot)$. Instead of sending to Bob a state from either one of the two ensembles, Alice can cheat by "entangling" her ancilla (system S_1) with Bob's system in the state Φ, and then measuring either one of the observables $\mathbb{A} = \{a_i\}$ and $\mathbb{B} = \{b_i\}$.

Notice that atomicity of identity occurs in QM, whereas it is not true in a classical probabilistic theory (see footnote 19). In classical mechanics one can gain information on the state without making a disturbance thanks to the non-atomicity of the identity transformation. According to Theorem 3 item (1) the need of disturbance for gaining information is a consequence of the purity of the preparationally faithful state, whence disturbance is the price to be paid for the reduction of the preparation complexity.

5.3.3 The Scalar product over effects induced by a symmetric faithful state

In this subsection I briefly review the construction in Ref. [17] of a scalar product over $\mathfrak{E}_{\mathbb{C}}$ via a symmetric faithful state, together with the corresponding operational definition of "transposed" and "complex conjugation" – with the composition of the two giving the adjoint.

According to Theorem 2 item (2), for two identical systems in finite dimensions any state that is preparationally faithful with respect to a system is also dynamically faithful with respect to the same system. Moreover, according to Postulate PFAITH, there always exists such a state, say Φ, which is symmetric under permutation of the two systems. The state Φ is then a symmetric real form over $\mathfrak{E}_{\mathbb{R}}$, whence it provides a non-degenerate scalar product over $\mathfrak{E}_{\mathbb{R}}$ via its Jordan form

$$\forall a, b \in \mathfrak{E}_{\mathbb{R}},\ {}_{\Phi}(b|a)_{\Phi} := |\Phi|(b, a) = \Phi(\varsigma(b), a), \tag{5.20}$$

where ς is the involution $\varsigma = \pi_+ - \pi_-$, $\pi_\pm$ denoting the orthogonal projectors over the positive (negative) eigenspaces of the symmetric form, or, explicitly, $\varsigma(a) := \sum_j \Phi(a, \tilde{f}_j) f_j$ and $\{\tilde{f}_j\}$ is the canonical Jordan basis.[37] Notice that the Jordan form is representation-dependent – i.e., it is defined through the reference test $\mathbb{L} = \{l_i\}$ – whereas its signature – i.e., the difference between the numbers of positive and negative eigenvalues – will be a property of the system S, and will generally depend on the specific probabilistic theory. For transformations $\mathscr{T} \in \mathfrak{T}_{\mathbb{R}}$ we define $a \circ \varsigma(\mathscr{T}) := \varsigma(\varsigma(a) \circ \mathscr{T}) =: a \circ \mathscr{Z} \circ \mathscr{T} \circ \mathscr{Z}$. For the identity transformation we have $\varsigma(\mathscr{I}) = \mathscr{Z} \circ \mathscr{Z} = \mathscr{I}$. Corresponding to a symmetric faithful bipartite state Φ one has the generalized transformation $\mathscr{T}_\Phi$, given by

$$a \circ \mathscr{T}_\Phi := \sum_k \Phi(l_k, a) l_k, \tag{5.21}$$

for a fixed orthonormal basis $\mathbb{L} = \{l_j\}$, and in terms of the corresponding symmetric scalar product $(\cdot, \cdot)_{\mathbb{L}}$ introduced in Section 5.2.10, one has

$$(a, b \circ \mathscr{T}_\Phi)_{\mathbb{L}} = (a \circ \mathscr{T}_\Phi, b)_{\mathbb{L}} = \Phi(a, b). \tag{5.22}$$

Using the dynamical and preparational faithfulness of Φ we have defined operationally the transposed $\mathscr{T}'$ of a transformation $\mathscr{T} \in \mathfrak{T}_{\mathbb{R}}$. Such an "operational" transposed is related to the transposed $\tilde{\mathscr{C}}$ under the scalar product $(\cdot, \cdot)_{\mathbb{L}}$ as $\mathscr{C}' = \mathscr{T}_\Phi \circ \tilde{\mathscr{C}} \circ \mathscr{T}_\Phi^{-1}$. It is easy to check that $\tilde{\mathscr{Z}} = \mathscr{Z} = \mathscr{Z}'$.

On the complex linear span $\mathfrak{T}_{\mathbb{C}}$ one can introduce a scalar product as the sesquilinear extension of the real symmetric scalar product $(\cdot, \cdot)_\Phi$ over $\mathfrak{E}_{\mathbb{R}}$ via

[37] In the diagonalizing orthonormal basis one has $s_j \delta_{ij} = \Phi(\tilde{f}_i, \tilde{f}_j) = |\lambda_j|^{-1} \Phi(f_i, f_j)$, $s_j = \pm 1$, $\tilde{f}_j = f_j / \sqrt{|\lambda_j|}$.

the complex conjugation $\eta(\mathscr{T}) = \mathscr{T}_{\mathrm{R}} - \mathrm{i}\mathscr{T}_{\mathrm{I}}$, $\mathscr{T}_{\mathrm{R,I}} \in \mathfrak{T}_{\mathbb{R}}$, and the adjoint for the sesquilinear scalar product is then given by

$$\mathscr{T}^{\dagger} = \mathscr{Z} \circ \eta(\mathscr{T}') \circ \mathscr{Z} = |\mathscr{T}_{\Phi}| \circ \eta(\tilde{\mathscr{T}}) \circ |\mathscr{T}_{\Phi}|^{-1}, \tag{5.23}$$

namely $\mathscr{T}^{\dagger} = \mathscr{Z} \circ \mathscr{T}' \circ \mathscr{Z}$ on real transformations $\mathscr{T} \in \mathfrak{T}_{\mathbb{R}}$. The Jordan involution ς thus plays the role of a complex conjugation on $\mathfrak{T}_{\mathbb{R}}$, which must be antilinearly extended to $\mathfrak{T}_{\mathbb{C}}$.

The faithful state Φ becomes a cyclic and separating vector of a GNS representation on noticing that $(\mathscr{A}^{(2)}\Phi)(\eta\varsigma b, a) = {}_{\Phi}(b, a \circ \mathscr{A})_{\Phi}$,[38] and in (5.23) one can recognize the Tomita–Takesaki modular operator of the representation [39].

5.4 Axiomatic interlude: exploring Postulates FAITHE and PURIFY

In this section we investigate two additional postulates of a probabilistic theory: Postulate FAITHE – the existence of a faithful effect (somehow dual to Postulate PFAITH) – and Postulate PURIFY – the existence of a purification for every state. As we will see, these new postulates bring the probabilistic theory closer and closer to QM. However, I was still unable to prove (or to find counterexamples) that with these two additional postulates the probabilistic theory *is* QM.

5.4.1 FAITHE: a postulate on a faithful effect

As previously mentioned, Postulate FAITHE is somehow the dual version of Postulate PFAITH.[39]

Postulate FAITHE (Existence of a faithful effect). *There exists a bipartite effect* $F \in \mathfrak{E}(\mathsf{S}^{\odot 2})$ *achieving the inverse of the isomorphism* $a \mapsto \omega_a := \Phi(a, \cdot)$. *More precisely,*

[38] The action of the algebra of generalized transformations on the first system corresponds to the transposed representation $(\mathscr{A}^{(1)}\Phi)(\eta\varsigma b, a) = \Phi(\eta\varsigma b \circ \mathscr{A}, a) = \Phi(\eta\varsigma b, a \circ \mathscr{A}') = (\mathscr{A}'^{(2)}\Phi)(\eta\varsigma b, a)$.

[39] At first sight it seems that the existence of an effect F such that $F_{23}\Phi_{12}\Phi_{34} = \alpha\Phi_{14}$ could be derived directly from PFAITH. Indeed, according to Lemma 5 for finite dimensions and identical systems we have $\mathfrak{S}_{\mathbb{F}}(\mathsf{S}^{\odot 2}) = \mathfrak{S}_{\mathbb{F}}(\mathsf{S})^{\otimes 2}$ and $\mathfrak{E}_{\mathbb{F}}(\mathsf{S}^{\odot 2}) = \mathfrak{E}_{\mathbb{F}}(\mathsf{S})^{\otimes 2}$ for $\mathbb{F} = \mathbb{R}, \mathbb{C}$. Moreover, according to Theorem 2 item (4) the map $a \mapsto \omega_a = \Phi(a, \cdot)$, for Φ symmetric preparationally faithful achieves the cone-isomorphism $\mathfrak{S}_+ \simeq \mathfrak{E}_+$, whence for the bipartite system one has $\mathfrak{S}_+(\mathsf{S}^{\odot 2}) \simeq \mathfrak{E}_+(\mathsf{S}^{\odot 2})$. This leads one to think that it should be possible to achieve a preparationally faithful state for $\mathsf{S}^{\odot 4}$ as the product $\Phi_{12}\Phi_{34}$. However, this is not necessarily true. In fact, since the map $\mathfrak{E}_{\mathbb{F}}(\mathsf{S})^{\otimes 2} \ni E \mapsto \Omega_E = E_{23}\Phi_{12}\Phi_{34}$ is a linear bijection between $\mathfrak{E}_{\mathbb{F}}(\mathsf{S})^{\otimes 2}$ and $\mathfrak{S}_{\mathbb{F}}(\mathsf{S})^{\otimes 2}$ (since $\mathsf{Span}_{\mathbb{F}}\{\Phi_{12}(\cdot, a)\Phi_{34}(b, \cdot) | a, b \in \mathfrak{E}\} = \mathfrak{S}_{\mathbb{F}}(\mathsf{S})^{\otimes 2} = \mathfrak{S}_{\mathbb{F}}(\mathsf{S}^{\odot 2})$) is cone-preserving, it sends separable effects to separable states, whence it sends non-separable effects to non-separable states (since it is one-to-one). However, it doesn't necessarily achieve the cone-isomorphism $\mathfrak{S}_+(\mathsf{S}^{\odot 2}) \simeq \mathfrak{E}_+(\mathsf{S}^{\odot 2})$, since it is not necessarily true that any bipartite state Ω is the mapped of a bipartite effect E_Ω (we remember that a cone-isomorphism is a bijection that preserves the cone *in both directions*). If by chance this were the case – i.e., $E \mapsto \Omega_E$ is a cone-isomorphism for $\mathsf{S}^{\odot 2}$ – then this would mean that there exists an effect $F \in \mathfrak{E}(\mathsf{S}^{\odot 2})$ such that $\Omega_F = \alpha\Phi$, with $0 < \alpha \leqslant 1$.

$$F_{23}(\omega_a)_2 = F_{23}\Phi_{12}(a,\cdot) = \alpha a_3, \quad 0 < \alpha \leqslant 1. \tag{5.24}$$

Notice that, since Φ establishes an isomorphism between the cones of states and effects, there must exist a generalized effect $F \in \mathfrak{E}_{\mathbb{R}}^{\otimes 2}$ satisfying (5.24), but we are not guaranteed that it is a physical one, i.e., $F \in \mathfrak{E}_{+}(\mathsf{S}^{\odot 2})$.

Let's denote by $\hat{F} = \alpha^{-1}F$ the rescaled effect in the cone. Equation (5.24) can be rewritten in different notation as follows:

$$\hat{F}(\omega_a,\cdot) = \hat{F}(\Phi(a,\cdot),\cdot) = a, \tag{5.25}$$

$$\Phi(a_\omega,\cdot) = \Phi(\hat{F}(\omega,\cdot),\cdot) = \omega. \tag{5.26}$$

(One needs to be careful with the notation in the multipartite case, e.g., in (5.26) $\Phi(\hat{F}(\omega,\cdot),\cdot) = \omega$ is actually a state, since $\hat{F}(\omega,\cdot)$ is an effect, etc.) Both faithful state Φ and faithful effect F can be used to express the state–effect pairing, namely

$$\zeta(b) = \Phi(a_\zeta, b) = \hat{F}(\omega_b,\zeta), \quad a_\zeta := \hat{F}(\zeta,\cdot),\ \omega_b := \Phi(b,\cdot), \tag{5.27}$$

or, substituting,

$$\zeta(b) = \Phi(\hat{F}(\zeta,\cdot), b) = \hat{F}(\Phi(b,\cdot),\zeta). \tag{5.28}$$

Equation (5.24) can also be rewritten as follows:

$$F_{23}\Phi_{12} = \alpha\,\mathbf{Swap}_{13}, \tag{5.29}$$

where $\mathbf{Swap}_{ij}$ denotes the transformation swapping S_i with S_j. In Figure 5.4 Postulate FAITHE is illustrated graphically.

Equation (5.29) means that by using the state Φ and the effect F one can achieve probabilistic **teleportation** of states from S_2 to S_4. In fact, one has

$$F_{23}\omega_2\Phi_{34} = F_{23}\Phi_{12}(a_\omega,\cdot)\Phi_{34} = \alpha\Phi_{14}(a_\omega,\cdot) = \alpha\omega_4. \tag{5.30}$$

Using the last identity we can also see that Postulate FAITHE is also equivalent to the identity

$$F_{23}\Phi_{12}\Phi_{34} = \alpha\Phi_{14}, \tag{5.31}$$

which by linearity is extended from local effects to all effects, by virtue of $\mathfrak{E}^{\odot 2} = \mathfrak{E}^{\otimes 2}$. With equivalent notation we can write $(\Phi,\Phi)(\cdot,F,\cdot) = \alpha\Phi$.

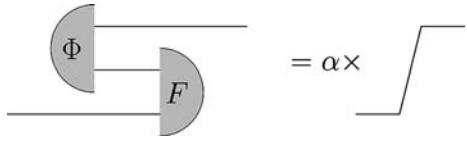


Fig. 5.4 An illustration of Postulate FAITHE.

The effect F is also completely faithful, in the sense that the correspondence $F_{\mathscr{A}} := F \circ (\mathscr{A}', \mathscr{I}) \Longleftrightarrow \mathscr{A}$ is bijective (in finite dimensions). In fact one has

$$[F \circ (\mathscr{A}', \mathscr{I})]_{23}(\Phi, \Phi) = \alpha(\mathscr{A}, \mathscr{I})\Phi, \tag{5.32}$$

and, since Φ is dynamically faithful (it is symmetric preparationally faithful), the correspondence $F_{\mathscr{A}} := F \circ (\mathscr{A}', \mathscr{I}) \Longleftrightarrow \mathscr{A}$ is one-to-one and surjective, whence it is a bijection (in finite dimensions). It is also easy to see that $F \circ (\mathscr{A}', \mathscr{I}) = F \circ (\mathscr{I}, \mathscr{A})$, since

$$\begin{aligned}[F \circ (\mathscr{I}, \mathscr{A})]_{23}(\Phi, \Phi) &= F_{23}(\Phi, (\mathscr{A}, \mathscr{I})\Phi) = F_{23}(\Phi, (\mathscr{I}, \mathscr{A}')\Phi) \\ &= \alpha(\mathscr{I}, \mathscr{A}')\Phi = \alpha(\mathscr{A}, \mathscr{I})\Phi = [F \circ (\mathscr{A}', \mathscr{I})]_{23}(\Phi, \Phi),\end{aligned} \tag{5.33}$$

whence transposition can be equivalently defined with respect to the faithful effect F. The bijection $F_{\mathscr{A}} := F \circ (\mathscr{I}, \mathscr{A}) \Longleftrightarrow \mathscr{A}$ is cone-preserving in both directions, since to every transformation there corresponds an effect, and to each effect $A \in \mathfrak{E}(\mathsf{S}^{\odot 2})$ there corresponds a transformation, since

$$A_{23}(\Phi, \Phi) = \Omega_A = (\mathscr{T}_{\Omega_A}, \mathscr{I})\Phi =: (\mathscr{T}_A, \mathscr{I})\Phi. \tag{5.34}$$

Therefore, the map $\mathscr{A} \mapsto F_{\mathscr{A}}$ realizes the cone-isomorphism $\mathfrak{E}_+(\mathsf{S}^{\odot 2}) \simeq \mathfrak{T}_+(\mathsf{S})$, which is just the composition of the weak self-duality and of the isomorphism $\mathfrak{S}_+(\mathsf{S}^{\odot 2}) \simeq \mathfrak{T}_+(\mathsf{S})$ due to PFAITH. However, as mentioned in footnote 39, the map

$$\mathfrak{E}_+(\mathsf{S}^{\odot 2}) \ni A \mapsto \Omega_A := A_{23}(\Phi, \Phi) \in \mathfrak{S}(\mathsf{S}^{\odot 2}) \tag{5.35}$$

is bijective between $\mathfrak{S}_{\mathbb{F}}(\mathsf{S}^{\odot 2})$ and $\mathfrak{E}_{\mathbb{F}}(\mathsf{S}^{\odot 2})$, but it does not realize the cone-isomorphism $\mathfrak{S}_+(\mathsf{S}^{\odot 2}) \simeq \mathfrak{E}_+(\mathsf{S}^{\odot 2})$, since it is not surjective over $\mathfrak{E}_+(\mathsf{S}^{\odot 2})$. Indeed, for $A \in \mathfrak{E}(\mathsf{S}^{\odot 2})$ physical effect, one has $A_{23}(\Phi, \Phi) = (\mathscr{T}_A, \mathscr{I})\Phi$ with $\mathscr{T}_A \in \mathfrak{T}(\mathsf{S})$ physical transformation. However, there is no guarantee that, vice versa, a physical transformation always has a corresponding physical effect, e.g., for the identity transformation in (5.31). It also follows that any bipartite observable $\mathbb{A} = \{A_l\}$ leads to the **totally depolarizing channel** $\mathscr{T}_{(e,e)}\omega = \chi$, $\forall \omega \in \mathfrak{S}$.[40] Using the faithfulness of F it is possible to achieve probabilistically any transformation on a state ω by performing a joint test on the system interacting with an ancilla, i.e., $(\omega\Phi)(F_{\mathscr{A}'}, \cdot) = \alpha\mathscr{A}\omega$ (for Stinespring-like dilations in an operational context see Ref. [31]).

[40] Indeed, one has $\sum_l (A_l)_{23}\omega_2\Phi_{34} = (e, e)_{23}\omega_2\Phi_{34} = \Phi_{12}(a_\omega, e)\Phi_{34}(e, \cdot) = \omega(e)\chi$.

More about the constant α. Notice that the number $0 < \alpha \leqslant 1$ is the probability of achieving teleportation $\alpha = (F_{23}\omega_2\Phi_{34})(e)$. It is independent of the state ω, and depends only on F, since it is given by $\alpha \equiv \alpha_F = [F_{23}\Phi_{12}\Phi_{34}](e, e)$. The maximum value maximized over all bipartite effects

$$\alpha(\mathsf{S}) = \max_{A\in\mathfrak{E}(\mathsf{S}^{\odot 2})} \{(\Phi, \Phi)(e, A, e)\} \tag{5.36}$$

is a property of the system S only, and depends on the particular probabilistic theory.

More on the relation between Postulates PFAITH and FAITHE. Postulate PFAITH guarantees the existence of a symmetric preparationally faithful state for each pair of identical systems $\mathsf{S}^{\odot 2}$. Now, consider the bipartite system $\mathsf{S}^{\odot 2} \odot \mathsf{S}^{\odot 2}$, and denote by $\mathbf{\Phi}$ a symmetric preparationally faithful state for it. The map $A \mapsto \Omega_A := \mathbf{\Phi}(A, \cdot, \cdot)\ \forall A \in \mathfrak{E}(\mathsf{S}^{\odot 2})$ establishes the state–effect cone-isomorphism for $\mathsf{S}^{\odot 2}$, whence there must exist an effect A_Φ such that

$$\mathbf{\Phi}(A_\Phi, \cdot, \cdot) = \beta\Phi, \quad 0 < \beta \leqslant 1. \tag{5.37}$$

Suppose now that the faithful state can be chosen in such a way that it maps separable states to separable effects as follows:

$$\mathbf{\Phi}(\cdot, \cdot, (a, b)) = \gamma(\omega_a, \omega_b) = \gamma\Phi(\cdot, a)\Phi(\cdot, b), \quad \gamma > 0. \tag{5.38}$$

Then one has

$$\gamma(A_\Phi)_{13}(\Phi, \Phi) = \mathbf{\Phi}(A_\Phi, \cdot, \cdot) = \beta\Phi, \tag{5.39}$$

namely, according to (5.31) one has $\beta^{-1}\gamma A_\Phi \equiv \hat{F}$, which is the effect whose existence is postulated by FAITHE. Notice, however, that the factorization (5.38) doesn't need to be satisfied. In other words, the automorphism relating the cone-isomorphism induced by $\mathbf{\Phi}$ to another cone-isomorphism that preserves local effects may be unphysical (see also footnote 39). One can instead require a stronger version of postulate PFAITH, postulating the existence of a preparationally **superfaithful** symmetric state Φ, also achieving a four-partite preparationally symmetric faithful state $\mathbf{\Phi}$ as $(\Phi, \Phi) = \mathbf{\Phi}$. A weaker version of such a postulate is thoroughly analyzed in Ref. [31], where it is also shown that it leads to Stinespring-like dilations of deterministic transformations.

The case of QM. It is a useful exercise to see how the present framework translates into the quantum case, and find which additional constraints can arise from a specific probabilistic theory. For simplicity we consider a maximally entangled state (with all positive amplitudes in a fixed basis) as a preparationally symmetric state Φ. The corresponding marginal state is given by the density matrix $d^{-1}I$, I

denoting the identity on the Hilbert space. For the constant α one has $\alpha = d^{-2}$, where d is the dimension of the Hilbert space. A simple calculation shows that the identity $\omega_a = \mathscr{T}_a'\chi$ for $\mathscr{T}_a \in a$ translates to[41]

$$\omega_a = \sqrt{\alpha}\varsigma(a), \qquad \Leftarrow \text{in QM}, \tag{5.40}$$

where the involution ς of the Jordan form in (5.20) here is also an automorphism of states/effects, whence identity (5.40) expresses the self-duality of QM. On rewriting (5.40) in terms of the faithful effect F (which would be an element of a Bell measurement), one obtains[42]

$$(\cdot, F)(\Phi, \cdot) = \sqrt{\alpha}|\Phi|, \qquad \Leftarrow \text{in QM}. \tag{5.41}$$

Another feature of QM is that the preparationally faithful symmetric state Φ is super-faithful, namely $\mathbf{\Phi} = (\Phi, \Phi)$ is preparationally faithful for $\mathsf{S}^{\odot 4}$.

5.4.2 PURIFY: a postulate on purifiability of all states

In the present section for completeness I briefly explore the consequences of assuming purifiability for all states, namely the following postulate.

Postulate PURIFY (Purifiability of states). *For every state ω of S there exists a pure bipartite state Ω of $\mathsf{S}^{\odot 2}$ having it as marginal state, namely*

$$\forall \omega \in \mathfrak{S}(\mathsf{S}),\ \exists \Omega \in \mathfrak{S}(\mathsf{S}^{\odot 2}) \text{ pure, such that } \Omega(e, \cdot) = \omega. \tag{5.42}$$

Postulate PURIFY has been analyzed in Ref. [31], where the following lemma is proved.

Lemma 6. *If Postulate PFAITH holds, then Postulate PURIFY implies the following assertions.*

(1) *Even without assuming purity of the preparationally faithful state Φ, the identity transformation is atomic, and purity of Φ can be derived.*
(2) $\mathfrak{S}_+ \equiv \mathsf{Erays}(\mathfrak{T}_+)\chi$*, i.e., each state can be obtained by applying an atomic transformation to the marginal state $\chi := \Phi(e, \cdot)$.*
(3) $\mathfrak{E}_+ \equiv e \circ \mathsf{Erays}(\mathfrak{T}_+)$*, i.e., each effect can be achieved with an atomic transformation.*

Points (2) and (3) correspond to the square root of states and effects in the quantum case.

[41] For $\Phi = d^{-1}\sum_{nm}|n\rangle|n\rangle\langle m|\langle m|$ the marginal state is $\chi = d^{-1}I$ and the Jordan involution is the complex conjugation with respect to the orthonormal basis $\{|n\rangle\}$. For quantum operation $\mathscr{T} = \sum_n T_n \cdot T_n^\dagger$ with corresponding effect $a = \sum_n T_n^\dagger T_n$, one has $\mathscr{T}'\chi = d^{-1}\sum_n T_n\, {}^t T_n^* = d^{-1}\sum_n (T_n^\dagger T_n)^* = \sqrt{\alpha}\varsigma(a)$.

[42] In fact, one has $\omega_a := \Phi(a, \cdot) = \sqrt{\alpha}\varsigma(a)$, namely $\Phi(\varsigma(a), \cdot) = \sqrt{\alpha}a$, i.e., $|\Phi|(a, \cdot) = \sqrt{\alpha}a$, and, using (5.25), one has $\sqrt{\alpha}\hat{F}(\Phi(a, \cdot), \cdot) = |\Phi|(a, \cdot)$, namely the statement.

5.5 What is special about quantum mechanics as a probabilistic theory?

The mathematical representation of the operational probabilistic framework derived up to now is completely general for any fair operational framework that allows local tests, test-calibration, and state preparation. These include not only QM and classical–quantum hybrid, but also other non-signaling non-local probabilistic theories such as the *PR-boxes* theories [20]. Postulate PFAITH has proved to be remarkably powerful, implying (1) the local observability principle, (2) the tensor-product structure for the linear spaces of states and effects, (3) weak self-duality, (4) realization of all states as transformations of the marginal faithful state $\Phi(e,\cdot)$, (5) locally indistinguishable ensembles of states corresponding to local observables – i.e., EPR-cheating in bit commitment – and more. By adding FAITHE one even has teleportation! However, despite all these positive landmarks, it is still unclear whether one can derive QM from these principles only.

What is then special about QM? The peculiarity of QM among probabilistic operational theories is the following.

Effects can not only be linearly combined, but also can be composed of each other, so that complex effects make a C^*-algebra.

Operationally the last assertion is odd, since *the notion of effect abhors composition!* Therefore, the composition of effects (i.e., the fact that they make a C^*-algebra, i.e., an operator algebra over complex Hilbert spaces) must be derived from additional postulates. What I will show here is the following.

With a single mathematical postulate, and assuming atomicity of evolution, one can derive the composition of effects in terms of composition of atomic events.

One thus is left with the problem of translating the remaining mathematical postulate into an operational one. Let's now examine the two postulates.

Postulate AE (Atomicity of evolution). *The composition of atomic transformations is atomic.*

This postulate is so natural that it looks obvious.[43] However, even though for atomic events $\mathscr{A}$ and $\mathscr{B}$ the event $\mathscr{C} = \mathscr{B} \circ \mathscr{A}$ is not refinable in the corresponding cascade-test, there is no guarantee that $\mathscr{C}$ is not refinable in any other test. We remember that mathematically atomic events belong to $\mathsf{Erays}(\mathfrak{T}_+)$, the extremal rays of the cone of transformations.

We now state the mathematical postulate.

[43] Indeed, when joining events $\mathscr{A}$ and $\mathscr{B}$ into the event $\mathscr{A} \wedge \mathscr{B}$, the latter is atomic if both $\mathscr{A}$ and $\mathscr{B}$ are atomic.

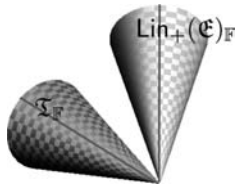


Fig. 5.5 The Choi–Jamiolkowski isomorphism between the cone $\mathfrak{T}_+$ of physical transformations and the cone $\mathsf{Lin}_+(\mathfrak{E}_{\mathbb{C}})$ of positive matrices over complex effects establishes a one-to-one correspondence between extremal-ray points of the two cones, identifying effects (modulo a phase) with atomic transformations (the lines over the cones represent a pair of corresponding rays).

Mathematical Postulate CJ (Choi–Jamiolkowski isomorphism (Figure 5.5)). *The cone of transformations is isomorphic*[44] *to the cone of positive bilinear forms over complex effects [27, 28], i.e.,* $\mathfrak{T}_+ \simeq \mathsf{Lin}_+(\mathfrak{E}_{\mathbb{C}})$.

In terms of a sesquilinear scalar product over complex effects, positive bilinear forms can be regarded as positive matrices over complex effects, i.e., elements of the cone $\mathsf{Lin}_+(\mathfrak{E}_{\mathbb{C}})$.

The extremal rays $\mathsf{Erays}(\mathsf{Lin}_+(\mathfrak{E}_{\mathbb{C}}))$ are rank-one positive operators $|x\rangle\langle x| \in \mathsf{Erays}(\mathsf{Lin}_+(\mathfrak{E}_{\mathbb{C}}))$ with $x \in \mathfrak{E}_{\mathbb{C}}$, and the map $\pi : x \mapsto \pi(x) := |x\rangle\langle x|$ is surjective over $\mathsf{Erays}(\mathsf{Lin}_+(\mathfrak{E}_{\mathbb{C}}))$. One has $\pi(x\mathrm{e}^{\mathrm{i}\phi}) = \pi(x)$, and $\pi^{-1}(|x\rangle\langle x|) = \{\mathrm{e}^{\mathrm{i}\phi}x\} \subseteq \mathfrak{E}_{\mathbb{C}}$, i.e., the set of complex effects mapped to the same rank-one positive operator is the set of complex effects that differ only by a multiplicative phase factor. We will denote by $|x| \in \mathfrak{E}_{\mathbb{C}}$ a fixed choice of representative for such an equivalence class,[45] introduce the phase corresponding to such a choice as $x =: |x|\mathrm{e}^{\mathrm{i}\phi(x)}$, and denote by $\mathfrak{E}_{\mathbb{C}}/\phi$ the set of equivalence classes, or, equivalently, of their representatives. Now, since the representatives $|x| \in \mathfrak{E}_{\mathbb{C}}/\phi$ are in one-to-one correspondence with the points on $\mathsf{Erays}(\mathsf{Lin}_+(\mathfrak{E}_{\mathbb{C}}))$, the CJ isomorphism establishes a bijective map between $\mathfrak{E}_{\mathbb{C}}/\phi$ and $\mathsf{Erays}(\mathfrak{T}_+)$ as follows:

$$\tau\colon \mathfrak{E}_{\mathbb{C}}/\phi \ni |x| \leftrightarrow \tau(|x|) \in \mathsf{Erays}(\mathfrak{T}_+). \tag{5.43}$$

5.5.1 Building up an associative algebra structure for complex effects

Assuming Postulate AE, we can introduce an associative composition between the effects in $\mathfrak{E}_{\mathbb{C}}/\phi$ via the bijection τ,

$$|a||b| := \tau^{-1}(\tau(|a|) \circ \tau(|b|)). \tag{5.44}$$

[44] For the definition of cone-isomorphisms, see footnote 34.

[45] An example of choice of representative is given by $||x|\rangle := \langle e_{\iota(x)}|\pi(x)|e_{\iota(x)}\rangle^{-1/2}\pi(x)|e_{\iota(x)}\rangle$, namely $|x| := |(x, e_{\iota(x)})|^{-1}(x, e_{\iota(x)})x$, with $\iota(x) = \min\{i\colon (x, e_i) \neq 0\}$, for given fixed basis for $\mathfrak{E}_{\mathbb{C}}$.

Notice that, by definition, $|a||b|$ is a representative of an equivalence class in $\mathfrak{E}_{\mathbb{C}}$, whence $|(|a||b|)| = |a||b|$. The above composition is extended to all elements of $\mathfrak{E}_{\mathbb{C}}$ by taking

$$ab := |a||b|\mathrm{e}^{\mathrm{i}\phi(a)}\mathrm{e}^{\mathrm{i}\phi(b)}, \tag{5.45}$$

and, since $|(|a||b|)| = |a||b|$, one has $|ab| = |a||b|$, and $\phi(ab) = \phi(a) + \phi(b)$. It follows that the extension is itself associative, since

$$\begin{aligned}(ab)c &= |ab||c|\mathrm{e}^{\mathrm{i}\phi(ab)+\mathrm{i}\phi(c)} = |a||b||c|\mathrm{e}^{\mathrm{i}\phi(a)+\mathrm{i}\phi(b)+\mathrm{i}\phi(c)}\\ &= |a||bc|\mathrm{e}^{\mathrm{i}\phi(a)+\mathrm{i}\phi(bc)} = a(bc).\end{aligned} \tag{5.46}$$

The composition is also distributive with respect to the sum, since it follows the same rules as those of complex numbers. We will denote by ι the identity in $\mathfrak{E}_{\mathbb{C}}/\phi$ when it exists, which also works as an identity for multiplication of effects as in (5.45). Notice that, since the identity transformation $\mathscr{I}$ is atomic, one has $\iota := \tau^{-1}(\mathscr{I}) \in \mathfrak{E}_{\mathbb{C}}/\phi$ according to (5.44).

5.5.2 *Building up a C^*-algebra structure over complex effects*

We want now to introduce a notion of adjoint for effects. We will do this in two steps: (a) we introduce an antilinear involution on the linear space $\mathfrak{E}_{\mathbb{C}}$; (b) we extend the associative product (5.45) under such antilinear involution.

(a) First we notice that the complex space $\mathfrak{E}_{\mathbb{C}}$ has been constructed as $\mathfrak{E}_{\mathbb{C}} = \mathfrak{E}_{\mathbb{R}} \oplus i\mathfrak{E}_{\mathbb{R}}$ starting from real combinations of physical effects $\mathfrak{E}_{\mathbb{R}} = \mathsf{Span}_{\mathbb{R}}(\mathfrak{E}_+)$, i.e., one has the unique Cartesian decomposition $x = x_{\mathrm{R}} + \mathrm{i}x_{\mathrm{I}}$ of $x \in \mathfrak{E}_{\mathbb{C}}$ in terms of $x_{\mathrm{R}}, x_{\mathrm{I}} \in \mathfrak{E}_{\mathbb{R}}$. We can then define the antilinear *dagger* involution $\dagger$ on $\mathfrak{E}_{\mathbb{C}}$ by taking $x^\dagger = x\ \forall x \in \mathfrak{E}_{\mathbb{R}}$ and $x^\dagger := x_{\mathrm{R}} - \mathrm{i}x_{\mathrm{I}}\ \forall x \in \mathfrak{E}_{\mathbb{C}}$. Notice that $\mathfrak{E}_{\mathbb{C}}$ is closed under such involution. On taking the involution of the defining identity $x =: |x|\mathrm{e}^{\mathrm{i}\phi(x)}$ one has $|x^\dagger| = |x|^\dagger \mathrm{e}^{-\mathrm{i}\phi(x^\dagger)-\mathrm{i}\phi(x)}$, which is consistently satisfied by choosing $|x^\dagger| = |x|^\dagger$ and $\phi(x^\dagger) = -\phi(x)\ \forall x \in \mathfrak{E}_{\mathbb{C}}$ (these identities are satisfied, e.g., for the choice of representative in footnote 45).

(b) The multiplications $a^\dagger b$ and $ab^\dagger$ are defined via the scalar product over $\mathfrak{E}_{\mathbb{C}}$ as follows:[46]

$$\forall c \in \mathfrak{E}_{\mathbb{C}}: \quad (c, a^\dagger b) := (ac, b), \qquad (c, ab^\dagger) := (cb, a). \tag{5.47}$$

This is possible since the scalar product over $\mathfrak{E}_{\mathbb{C}}$ is supposed to be non-degenerate. It is then easy to verify that one has the identities $(ab)^\dagger = b^\dagger a^\dagger$ and $\iota^\dagger = \iota$.

In this way $\mathfrak{E}_{\mathbb{C}}$ is closed under complex linear combinations, the adjoint, and associative composition, and possibly contains the identity element ι; that is, it is an

[46] The right and left multiplications are just special elements of the algebra $\mathsf{Lin}(\mathfrak{E}_{\mathbb{C}})$, whence their adjoints are definable via the scalar product as usual.

associative complex algebra with adjoint, closed with respect to the adjoint. The scalar product on $\mathfrak{E}_{\mathbb{C}}$ in conjunction with the identity leads to a strictly positive linear form over $\mathfrak{E}_{\mathbb{C}}$, defined as $\Phi = (\iota, \cdot)$, and one has $\Phi(a^\dagger b) = (\iota, a^\dagger b) = (a, b)$.[47] Such a form is also a *trace*, i.e., it satisfies the identity $\Phi(ba) = \Phi(ab)$, which can be easily verified using definitions (5.47).[48] The complex linear space of the algebra closed with respect to the norm induced by the scalar product makes it a Hilbert space, and the action of the algebra over itself regarded as a Hilbert space makes it an operator algebra.[49] It is a standard result of the theory of operator algebras that the closure of $\mathfrak{E}_{\mathbb{C}}$ under the operator norm (which is guaranteed in finite dimensions) is a C^*-algebra. We have therefore built a C^*-algebra structure over the complex linear space of effects $\mathfrak{E}_{\mathbb{C}}$. This is the *cyclic representation* [39] given by

$$\Phi(a) = \langle \iota | \pi_\Phi(a) | \iota \rangle, \tag{5.48}$$

π_Φ denoting the algebra representation corresponding to Φ.[50] In our case one has $\pi_\Phi(a)|\iota\rangle = |a\rangle$, along with the *trace* property $\langle \iota | \pi_\Phi(a)\pi_\Phi(b) | \iota \rangle = \langle \iota | \pi_\Phi(b)\pi_\Phi(a) | \iota \rangle$. The latter can be actually realized as a trace as $\Phi(a^\dagger b) = \mathrm{Tr}\,[O(a)^\dagger O(b)]$, via a faithful representation $O\colon a \mapsto O(a) \in \mathsf{Lin}(\mathsf{H})$ of the algebra $\mathfrak{E}_{\mathbb{C}}$ as a sub-algebra of $\mathsf{Lin}(\mathsf{H})$ of operators over a Hilbert space H with dimension $\dim(\mathsf{H})^2 \geqslant \dim(\mathfrak{E}_{\mathbb{C}})$. In this way, one has $\pi_\Phi(a) = (O(a) \otimes I)$ with the cyclic vector represented as $|\iota\rangle = \sum_n |n\rangle \otimes |n\rangle$, $\{|n\rangle\}$ being any orthonormal basis for H.

5.5.3 Recovering the action of transformations over effects

In order to complete the mathematical representation of the probabilistic theory, we now need to define the action of the elements of $\mathfrak{T}_{\mathbb{C}}$ over $\mathfrak{E}_{\mathbb{C}}$, and to select the cone of physical transformations $\mathfrak{T}_+$. We will show that $\mathfrak{T}_+$ is given by the completely positive linear maps on $\mathfrak{E}_{\mathbb{C}}$, namely the linear maps of the Kraus form, i.e., the atomic transformations act on $x \in \mathfrak{E}_{\mathbb{C}}$ as $x \circ \tau(|a|) = |a|^\dagger x |a| \equiv a^\dagger x a$.

First, notice that the full span $\mathsf{Lin}(\mathfrak{E}_{\mathbb{C}})$ is recovered from $\mathsf{Erays}(\mathsf{Lin}_+(\mathfrak{E}_{\mathbb{C}}))$ via the polarization identity

$$|a\rangle\langle b| = \frac{1}{4} \sum_{k=0}^{3} \mathrm{i}^k |(a + \mathrm{i}^k b)\rangle\langle(a + \mathrm{i}^k b)|. \tag{5.49}$$

[47] The form is strictly positive since $\Phi(a^\dagger a) = (a, a) \geqslant 0$, with the equals sign only if $a = 0$, since the scalar product is non-degenerate.

[48] One has $\Phi(ab) = (\iota, ab) = (\iota, a(b^\dagger)^\dagger) = (b^\dagger, a)$ and $\Phi(ba) = (\iota, ba) = (\iota, (b^\dagger)^\dagger a) = (b^\dagger, a)$.

[49] This construction is a special case of the Gelfand–Naimark–Segal (GNS) construction [40], in which the form Φ is a trace. In the standard GNS construction the form Φ may be degenerate, i.e., one can have $\Phi(a^\dagger a) = 0$ for some $a \neq 0$, and the vectors of the representation are built up as equivalence classes modulo vectors having $\Phi(a^\dagger a) = 0$.

[50] This means that $\pi_\Phi(a)\pi_\Phi(b) = \pi_\Phi(ab)$ and $\pi_\Phi(a^\dagger) = \pi_\Phi(a)^\dagger$.

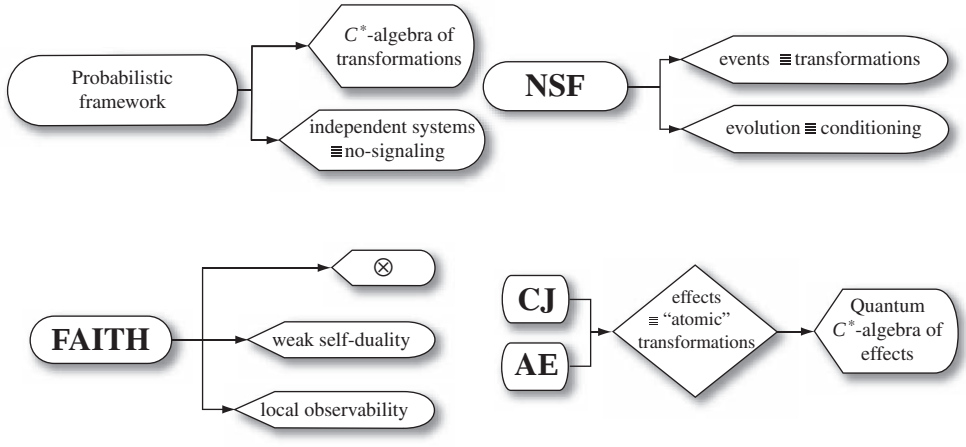


Fig. 5.6 An operational axiomatic framework for quantum mechanics: a summary of the relevant logical implications.

Correspondingly, we introduce the generalized transformations

$$\tau(b,a) := \frac{1}{4}\sum_{k=0}^{3} \mathrm{i}^k \tau(|a + \mathrm{i}^k b|) \in \mathfrak{T}_{\mathbb{C}}. \tag{5.50}$$

The map

$$|a\rangle\langle b| \mapsto \chi(|a\rangle\langle b|) := b^\dagger \cdot a \tag{5.51}$$

is a CJ isomorphism: it represents a bijective map between the cones $\mathsf{Lin}_+(\mathfrak{E}_{\mathbb{C}})$ and $\mathfrak{T}_+$, which can be extended to a cone-preserving linear bijection between $\mathsf{Lin}(\mathfrak{E}_{\mathbb{C}})$ and $\mathfrak{T}_{\mathbb{C}} \equiv \mathsf{Lin}(\mathfrak{E}_{\mathbb{C}})$.[51] As a consequence of (5.44), the CJ isomorphism $\tau : |a| \mapsto \tau(|a|)$ will differ from the isomorphism χ by an automorphism $\mathcal{U}$ of the C^*-algebra of effects; that is, one has $x \circ \tau(|a|) = \mathcal{U}(a^\dagger) x \mathcal{U}(a)$, with $\mathcal{U}(a) = u^\dagger a u$ with $uu^\dagger = u^\dagger u = \iota$. It follows that the probabilistic equivalence classes are given by $[\tau(|a|)]_{\mathrm{eff}} = e \circ \tau(|a|) = u^\dagger a^\dagger a u$. Notice that $[\tau(\iota)]_{\mathrm{eff}} = u^\dagger \iota^\dagger \iota u = \iota$; that is, ι coincides with the deterministic effect $\iota = e$. Complex effects are thus recovered from atomic transformations via the identity $e \circ \tau(e,a) = u^\dagger a u$. Figure 5.6 is a flow diagram summarizing the relevant logical implications of the present operational axiomatic framework for QM.

[51] This can be directly checked using the operator algebra representation built over $\mathfrak{E}_{\mathbb{C}}$, whereas the isomorphism corresponds to the map $O(b^\dagger x a) = \chi(|a\rangle\langle b|)(x) = \mathrm{Tr}_1[(O(x) \otimes I)|a\rangle\langle b|]$, and, reversely, $|a\rangle\langle b| = \chi^{-1}(\tau(b,a)) = (\tau(b,a) \otimes \mathscr{I})(|\iota\rangle\langle\iota|)$.

5.5.4 Reconstructing quantum mechanics from the probabilistic theory

It is now possible to reconstruct from the probability tables of the systems the full C^*-algebra of complex effects $\mathfrak{E}_{\mathbb{C}}$ as an operator algebra $\mathfrak{E}_{\mathbb{C}} \subseteq \oplus_i \mathsf{Lin}(\mathsf{H}_i)$. Here is the recipe.

(1) Look for all sub-cones $(\mathfrak{E}_+)_i$ invariant under $\mathfrak{T}_+$.
Then, for each i:
(2) introduce a complex Hilbert space H_i such that $(\mathfrak{E}_{\mathbb{C}})_i \subseteq \mathsf{Lin}(\mathsf{H}_i)$, i.e., with $\dim(\mathsf{H}_i) = \lceil \sqrt{\dim[(\mathfrak{E}_{\mathbb{C}})_i]} \rceil$, $\lceil x \rceil$ the smallest integer greater than x;
(3) represent e as the identity over $\oplus_i \mathsf{H}_i$;
(4) build $(\mathfrak{T}_{\mathbb{C}})_i \subseteq \mathsf{Lin}(\mathsf{Lin}(\mathsf{H}_i))$;
(5) look for atomic transformations $\mathsf{Erays}(\mathfrak{T}_+)_i$;
(6) for a given atomic transformation $\mathscr{A} \in \mathsf{Erays}(\mathfrak{T}_+)_i$ take an operator $A \in \mathsf{Lin}(\mathsf{H}_i)$ to represent $\mathscr{A}$ as $A^\dagger \cdot A \in \mathsf{Lin}(\mathsf{Lin}(\mathsf{H}_i))$;
(7) represent $[\mathscr{A}]_{\text{eff}}$ as $A^\dagger A$;
(8) repeat steps 6 and 7 for another transformation $\mathscr{B}$;
(9) compose $\mathscr{C} = \mathscr{B} \circ \mathscr{A}$ and represent $\mathscr{C}$ as $C^\dagger \cdot C$, with $C = AB$;
(10) repeat steps 8 and 9 to build the whole algebra of effects and the corresponding representation of the algebra of transformations; and
(11) construct states as density operators using the Gleason-like theorem [41] for effects [42, 43].

5.6 Conclusions

Theoretical physics should be, in essence, a mathematical "representation" of reality. By "representation" we mean describing one thing by means of another, to connect the object that we want to understand – the *thing-in-itself* – with an object that we already know well – the *standard*. In theoretical physics we lay down morphisms from structures of reality to corresponding mathematical structures: groups, algebras, vector spaces, etc., each mathematical structure capturing a different side of reality.

Quantum mechanics somehow goes differently. We have a beautiful simple mathematical structure – Hilbert spaces and operator algebras – with unprecedented predictive power in the entire physical domain. However, we don't have morphisms from the operational structure of reality into a mathematical structure. In this sense we can say that QM is not yet truly a "representation" of reality. A large part of the formal structure of QM is a set of formal tools for describing the process of gathering information in any experiment, independently of the particular physics involved. It is mainly a kind of information theory, a theory about our knowledge of physical entities rather than about the entities themselves. If we were to strip off this informational part from the theory, what

Table 5.1 *A summary of notation*

Symbol(s)	Meaning	Related quantities
$\mathsf{S}_1 \odot \mathsf{S}_2$	Bipartite system obtained by composing S_1 with S_2	
$\mathsf{S} = \{\mathbb{A}, \mathbb{B}, \mathbb{C}, \ldots\}$	System	
$\mathbb{A}, \mathbb{B}, \mathbb{C}, \ldots$	Tests	$\mathbb{A} = \{\mathscr{A}_j\}$, Test := set of possible events
$\mathscr{A}, \mathscr{B}, \mathscr{C}, \ldots$	Events $\equiv$ transformations	
ω	States, $\mathfrak{S}$ convex set of states	$\omega(\mathscr{A})$: probability that event $\mathscr{A}$ occurs in state ω
$\mathfrak{T}$	Convex monoid of transformations/events	$\mathfrak{T}_{\mathbb{R}}$,$\mathfrak{T}_{\mathbb{C}}$: linear spans of $\mathfrak{T}$, $\mathfrak{T}_+$: convex cone
$[\mathscr{A}]_{\text{eff}}$	Effect containing event $\mathscr{A}$	
$a, b, c, \ldots$	Effects	e: deterministic effect
$\mathfrak{E}$	Convex set of effects	$\mathfrak{E}_{\mathbb{R}}$, $\mathfrak{E}_{\mathbb{C}}$: linear spans of $\mathfrak{E}$, $\mathfrak{E}_+$: convex cone
$\mathbb{L} = \{l_j\}$	observable	$\sum_{l_i \in \mathbb{L}} l_i = e$
$\mathfrak{T}_{\mathbb{C}}$	C^*-algebra of transformations/events	
$a \circ \mathscr{T}$	Operation of transformation $\mathscr{T}$ over effect a	
$\omega_{\mathscr{A}}$	Conditioned states	$\omega_{\mathscr{A}} := \omega(\cdot \circ \mathscr{A})/\omega(\mathscr{A})$, $\mathscr{A}\omega = \omega(\cdot \circ \mathscr{A})$
$\mathsf{Lin}_+(\mathfrak{E}_{\mathbb{C}})$	Cone of linear maps corresponding to positive bilinear forms over $\mathfrak{E}_{\mathbb{C}}$	$\mathsf{Lin}_+(\mathfrak{E}_{\mathbb{C}}) = \{\mathscr{T} \in \mathfrak{T}_{\mathbb{C}} : (a, a \circ \mathscr{T}) \geqslant 0, \forall a \in \mathfrak{E}_{\mathbb{C}}\}$

would be left should be the true general principle from which QM should be derived.

In the present work I have analyzed the possibility of deriving QM as the mathematical representation of a fair operational framework made of a set of rules that allows one to make predictions about future events on the basis of suitable tests. The two postulates NSF and PFAITH need to be satisfied by an operational framework that is fair, the former in order for one to be able to make predictions that are based on present tests, the latter to allow calibrability of any test and preparability of any state. We have seen that all theories satisfying NSF admit a C^*-algebra representation of events as linear transformations of complex effects. On the basis of a very general notion of dynamical independence, all such theories are *non-signaling*. The C^*-algebra representation of events is just the informational part of the theory. We have then added Postulate PFAITH. Postulate PFAITH has been proved to be remarkably powerful, implying the local observability principle, the tensor-product structure for the linear spaces of states and effects, weak

self-duality, and a list of features such as realization of all states as transformations of the marginal faithful state $\Phi(e, \cdot)$, locally indistinguishable ensembles of states corresponding to local observables – i.e., EPR-cheating in bit commitment, and more. We have then explored a postulate dual to PFAITH, Postulate FAITHE for effects, thus deriving additional quantum features, such as teleportation. We feel that we are really close to QM: maybe we are already there and we only need to prove it! All the consequences of these postulates need to be explored further. I have also reported some consequences of a postulate about the purifiability of all states. In any case, we have seen that, whatever the missing postulate is, it must establish a one-to-one correspondence between complex effects and atomic transformations, which, assuming atomicity of evolution (Postulate AE) will make also effects a C^*-algebra. This is what is special about QM (and all hybrid quantum–classical theories), and will exclude other non-signaling probabilistic theories of the kind of the PR boxes.[52] We have seen that the correspondence between effects and atomic transformations is established by the Choi–Jamiolkowski isomorphism, which is hoped to be not too far from an operational principle.

The notation used is summarized in Table 5.1

Acknowledgments

The present research program started in August 2003 in Evanston IL at Northwestern University, where it continued during many following visits, far from the distractions of normal life, thanks to the generous hospitality of Horace Yuen. During these years the work has benefited from feedback and encouragement from many authors and colleagues at various conferences, where preliminary results were presented, most frequently at Växjö and Perimeter, but also at meetings in Losini, Zurich, Tokyo, Sendai, Cambridge (UK), Obergurgl, and Leiden. In particular the present text was initiated in Cambridge (UK) thanks to the hospitality of Robert Spekkens on the occasion of the workshop Operational Probabilistic Theories as Foils to Quantum Theory, July 2–13, 2007, and was finished in Nagoya thanks to the hospitality of Masanao Ozawa in March 2009. I acknowledge useful discussions with numerous interested colleagues, especially with Dirk Schlingeman, Tony Short, Ray Streater, Chris Fuchs, Marcus Appleby, Karl Svozil, Gregg Jaeger, and Lucien Hardy, and valuable encouragement at the very beginning from Maria Luisa dalla Chiara, Enrico Beltrametti, Guido Bacciagaluppi, and Jos Uffink. I'm most grateful to my talented disciples Giulio Chiribella and Paolo Perinotti for pointing out errors in the previous versions of this text, and for their helpful

[52] The PR boxes in principle satisfy NSF, and can admit a dynamical faithful state, e.g., the boxes of Ref. [44] (private discussion with Tony Short).

comments. The work has also greatly gained from a thorough discussion with Alexander Wilce, Howard Barnum, and Reinhard Werner held in Obergurgl during the QICS Workshop on Foundational Structures for Quantum Information and Computation, September 14–20, 2008. Finally, I'm indebted to Masanao Ozawa and Reinhard Werner for consistently supporting this research and following its progress, with deep analysis and very useful suggestions.

References

[1] G. Birkhoff and J. von Neumann, The logic of quantum mechanics, *Ann. Math.* **37**, 823 (1936).
[2] P. Jordan, J. von Neumann, and E. Wigner, On an algebraic generalization of the quantum mechanical formalism, *Ann. Math.* **35**, 29 (1934).
[3] V. S. Varadarajan, Probability in physics and a theorem on simultaneous observability, *Commun. Pure Appl. Math.* **15**, 189 (1962).
[4] G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (New York: Benjamin, 1963).
[5] C. Piron, *Foundations of Quantum Physics* (Massachusets: Benjamin, 1976).
[6] I. E. Segal, Postulates for general quantum mechanics, *Ann. Math.* **48**, 930 (1947).
[7] F. Strocchi, *An Introduction to the Mathematical Structure of Quantum Mechanics: A Short Course for Mathematicians* (Singapore: World Scientific, 2005).
[8] E. M. Alfsen and F. W. Shultz, State spaces of operator algebras: basic theory, orientations, and C^*-products, *Mathematics: Theory and Applications* (Boston, MA: Birkhäuser, 2001).
[9] E. M. Alfsen and F. W. Shultz, Geometry of state spaces of operator algebras, *Mathematics: Theory and Applications* (Boston, MA: Birkhäuser, 2002).
[10] W. Thirring, *Quantum Mathematical Physics* (Berlin: Springer, 2004).
[11] G. Ludwig, *An Axiomatic Basis for Quantum Mechanics I: Derivation of Hilbert Space Structure* (Berlin: Springer, 1985).
[12] L. Hardy, Quantum theory from five reasonable axioms, quant-ph/0101012v4 (2001).
[13] R. Clifton, J. Bub, and H. Halvorson, Characterizing quantum theory in terms of information-theoretic constraints, *Foundations Phys.* **33**, 1561 (2003).
[14] G. M. D'Ariano, Operational axioms for quantum mechanics, in *Foundations of Probability and Physics – 4*, G. Adenier, C. A. Fuchs, and A. Yu. Khrennikov (eds.) (Melville, NY: American Institute of Physics, 2007), p. 79.
[15] G. M. D'Ariano, On the missing axiom of quantum mechanics, in *Quantum Theory, Reconsideration of Foundations – 3*, G. Adenier, A. Yu. Khrennikov, and T. M. Nieuwenhuizen (eds.) (Melville, NY: American Institute of Physics, 2006), p. 114.
[16] G. M. D'Ariano, How to derive the Hilbert-space formulation of quantum mechanics from purely operational axioms, in *Quantum Mechanics*, A. Bassi, D. Duerr, T. Weber, and N. Zanghi (eds.) (Melville, NY: American Institute of Physics, 2006), p. 101.
[17] G. M. D'Ariano, Operational axioms for C^*-algebra representation of transformations, in *Quantum Theory, Reconsideration of Foundations 4*, G. Adenier, A. Yu. Khrennikov, P. Lahti, V. I. Man'ko, and Th. M. Nieuwenhuizen (eds.) (Melville, NY: American Institute of Physics, 2007), p. 44.

[18] G. M. D'Ariano, Operational axioms for a C^*-algebraic formulation for quantum mechanics, in *Proceedings of the 8th International Conference on Quantum Communication, Measurement and Computing*, O. Hirota, J. H. Shapiro, and M. Sasaki (eds.) (Tokyo: NICT Press, 2007), p. 345.
[19] G. M. D'Ariano, Where the mathematical structure of quantum mechanics comes from, in *Beyond the Quantum*, T. M. Nieuwenhuizen, V. Spicka, B. Mehmani, M. J. Aghdami, and A. Yu. Khrennikov (eds.) (Singapore: World Scientific, 2007), p. 191.
[20] S. Popescu and D. Rohrlich, Quantum non-locality as an axiom, *Foundations Phys.* **24**, 379 (1994).
[21] B. S. Cirel'son, Quantum generalizations of Bell's inequality, *Lett. Math. Phys.* **4**, 93 (1980).
[22] C. Simon, V. Bužek, and N. Gisin, No-signaling condition and quantum dynamics, *Phys. Rev. Lett.* **87**, 170405 (2001).
[23] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, Generalized no-broadcasting theorem, *Phys. Rev. Lett.* **99**, 240501 (2007).
[24] J. Barrett, Information processing in generalized probabilistic theories, quant-ph/0508211 (2005).
[25] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, Teleportation in general probabilistic theories, quant-ph/0805.3553 (2008).
[26] H. Barnum, Coordinating quantum agents' perspectives: convex operational theories, quantum information, and quantum foundations, quant-ph/0611110 (2006).
[27] M.-D. Choi, Completely positive linear maps on complex matrices, *Lin. Alg. Appl.* **10**, 285 (1975).
[28] A. Jamiolkowski, Linear transformations which preserve trace and positive semidefiniteness of operators, *Rep. Math. Phys.* **3**, 275 (1972).
[29] A. Rényi, *Foundations of Probability* (Mineola, NY: Dover, 2007).
[30] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Quantum circuit architecture, *Phys. Rev. Lett.* **101**, 060401 (2008).
[31] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Reversible realization of physical processes in probabilistic theories, arXiv:0908.1583 (2009).
[32] A. Barvinok, *A Course in Convexity* (Providence, RI: American Mathematical Society, 2002).
[33] J. Béllissard and B. Jochum, Homogeneous self-dual cones, versus Jordan algebras. The theory revisited, *Ann. Inst. Fourier (Grenoble)* **28**, 27 (1978).
[34] C. A. Fuchs, Quantum Mechanics as Quantum Information (and only a little more), quant-ph/0205039 (2002).
[35] D. Koehler and P. Rosenthal, On isometries of normed linear spaces, *St. Math.* **36**, 213 (1970).
[36] G. M. D'Ariano, No-signalling, dynamical independence and the local observability principle, *J. Phys. A* **40**, 8137 (2006).
[37] V. B. Scholz and R. F. Werner, Tsirelson's problem, arXiv 0812.4305 (2008).
[38] G. M. D'Ariano, D. Kretschmann, D. Schlingemann, and R. F. Werner, Reexamination of quantum bit commitment: the possible and the impossible, *Phys. Rev. A* **76**, 27 (2007).
[39] R. Haag, *Local Quantum Physics: Fields, Particles, Algebras* (Berlin: Springer, 1996).
[40] I. M. Gelfand and M. A. Neumark, On the imbedding of normed rings into the ring of operators in Hilbert space, *Mat. Sb.* **12**, 197 (1943).
[41] A. M. Gleason, Measures on the closed subspaces of a Hilbert space, *J. Math. Mech.* **6**, 885 (1957).

[42] P. Busch, Quantum states and generalized observables: a simple proof of Gleason's theorem, *Phys. Rev. Lett.* **91**, 120403 (2003).
[43] C. M. Caves, C. A. Fuchs, K. Manne, and J. Renes, *Gleason's Theorem with POVMS* (unpublished, 2000).
[44] A. J. Short, S. Popescu, and N. Gisin, Entanglement swapping for generalized non-local correlations, *Phys. Rev. A* **73**, 012101 (2006).

6

What probabilities tell about quantum systems, with application to entropy and entanglement

John M. Myers and F. Hadi Madjid

6.1 Introduction

The use of parameters to describe an experimenter's control over the devices used in an experiment is familiar in quantum physics, for example in connection with Bell inequalities. Parameters are also interesting in a different but related context, as we noticed when we proved a formal separation in quantum mechanics between linear operators and the probabilities that these operators generate. In comparing an experiment against its description by a density operator and detection operators, one compares tallies of experimental outcomes against the probabilities generated by the operators but not directly against the operators. Recognizing that the accessibility of operators to experimental tests is only indirect, via probabilities, motivates us to ask what probabilities tell us about operators, or, put more precisely, "what combinations of a parameterized density operator and parameterized detection operators generate any given set of parametrized probabilities?"

Here, we review and augment recent proofs that any given parameterized probabilities can be generated in very diverse ways, so that a parameterized probability measure, detached from any of the (infinitely many) parameterized operators that generate it, becomes an interesting object in its own right. By detaching a parameterized probability measure from the operators that may have led us to it, we (1) strengthen Holevo's bound on a quantum communication channel and (2) clarify a role for multiple levels of modeling in an example based on quantum key distribution. We then inquire into some parameterized probability measures generated by entangled states and into the topology of the associated parameter spaces; in particular we display some previously overlooked topological features of level sets of these probability measures.

As described quantum mechanically, an actual or contemplated experimental trial parses into (1) "a preparation" expressed by a density operator (for a pure or mixed quantum state) and (2) "a measurement" expressed by the detection

Philosophy of Quantum Information and Entanglement, ed. A. Bokulich and G. Jaeger. Published by Cambridge University Press.

operators associated with a positive operator-valued measure (POVM) [1–3]. By a convention of quantum physics, one combines a density operator ρ and a detection operator $M(\omega)$ in the "trace" formula by which these two generate a probability $\mu(\omega)$ for a measurable event ω:

$$\mu(\omega) = \mathrm{Tr}\,[\rho M(\omega)]. \tag{6.1}$$

(A special case is that of pure states (wavefunctions) $|\psi\rangle$ for which the density operator has the form $\rho = |\psi\rangle\langle\psi|$; on the POVM side, a special case is that in which the detection operators are projections.)

In comparing an experiment against its description by a density operator and detection operators, one compares tallies of experimental outcomes [4] not against the operators directly but against the probabilities calculated from the density operator and the detection operators per (6.1). Recognizing that the operators are tested experimentally only indirectly by way of probabilities, we wondered what can given probabilities tell us about density operators and detection operators? (To begin with, we ask this question without the "tensor-product" restrictions imposed on the detection operators in some prior investigations [5].) A little thought shows that any given probability $\mu(\omega)$ can be generated by distinct choices of ρ and $M(\omega)$, but we wanted to know how the dependence of μ, ρ, and M on experimental control affects this freedom of choice. Characterizing control over an experiment by some list k of parameter values, we introduced a set $\mathcal{K}$ of lists of parameter values as the domain of three functions:

(1) a function that assigns to each parameter list $k \in \mathcal{K}$ a density operator $\rho^{(k)}$;
(2) a function that assigns to each k a POVM $M^{(k)}$; and
(3) a function that assigns to each k a probability measure $\mu^{(k)}$.

In partial answer to our question, we proved that no assignment of probability measures to parameter lists can ever uniquely determine the density operators and POVMs [6]. This proof reveals a logical gap between parameterized probabilities and the parameterized operators that generate them, thereby making parameterized probability measures objects of interest in their own right, as will be demonstrated in the following sections. In Sections 6.3.1 and 6.3.2, we give two examples of uses of parameterized probability measures independently of the particular quantum states and POVMs that led us to them. The first example pertains to the information capacity of a quantum channel, while the second pertains to quantum cryptography.

In Section 6.4 we study parameterized probability measures associated with some entangled states. While pure entangled states violate Bell inequalities [7–9], these inequalities pertain directly not to quantum states as density operators but to the parameterized probability measures ensuing from these states conjoined to

parameterized POVMs. We inquire into the topology of the associated parameter spaces and we display topological features of level sets of the probability measures as functions on the parameter space. Some open questions are sketched. In the next section we begin with definitions, some probably novel, others previously known but perhaps unfamiliar.

6.2 Parameter spaces and quantum models

In setting up an experiment, say on an optics bench, one reaches not for a density operator or a POVM, but for lenses and lasers. Yet neither "lens" nor "laser" is a term in the language of quantum mechanics. Indeed, if the devices of an experiment – lasers, detectors, filters, and so on – were fixed, a quantum theorist would abstract the devices used in the experiment into a single point to which he or she attaches a density operator and a POVM. Generally, however, the experimenter adjusts the devices: by turning one knob to rotate a polarizing filter, another knob to change the position of a lens, another to control the temperature of a crystal, etc. Correspondingly, the experiment tallies detections not just for a run of trials corresponding to a single value of a parameter list, but for runs corresponding to a variety of values. We express the possible positions of the "knobs" for some particular experiment, actual or contemplated, by a set $\mathcal{K}$; for some fixed positive integer ℓ, any element $k \in \mathcal{K}$ is a list $\{k_1, k_2, \ldots, k_\ell\}$ in which k_j can be either a continuous (real-valued) parameter or a discrete parameter. In most cases there is a topology on $\mathcal{K}$, in which case we speak of a parameter *space*. We will see an example in which $\mathcal{K}$ is a torus. For the sake of simplicity, here we restrict attention to experimental parameters that, while they can change in value from trial to trial, are held constant during any one trial. (In contrast, Berry phases arise from modeling what happens to a quantum state when parameters are changed during a trial [10].)

An experimenter can have occasion to add a device or to replace a device with one that works over a wider range of parameter values. Such an extension of the experiment calls for extending its description, including extending the set $\mathcal{K}$ into a bigger set $\mathcal{K}'$ via one or both of two types of injective mapping $\Xi\colon \mathcal{K} \hookrightarrow \mathcal{K}'$:

(1) $\mathcal{K}'$ increases the range of one or more of the parameters of the list in $\mathcal{K}$, so that $\mathcal{K} \subset \mathcal{K}'$; and/or

(2) the lists in $\mathcal{K}'$ contain more parameters, so that, for some extra parameters in a set $\mathcal{K}''$, we have $\mathcal{K}' = \mathcal{K} \times \mathcal{K}''$. (Any parameter list $k' \in \mathcal{K}'$ has the form $k' = k \| k''$, where $k \in \mathcal{K}$, $k'' \in \mathcal{K}''$ and "$\|$" indicates concatenation.) In this type of extension of a parameter set, the injection can immerse (or even embed) a lower-dimensional manifold into one of higher dimension.

Rather than extending it, an experimenter sometimes simplifies an experiment, which leads a theorist to simplify a parameter space, for instance by making some of the parameters on the list dependent on others. The simplest case is to fix the values of certain parameters of the list; then a simplified parameter set $\mathcal{K}'$ comes from deleting the parameters whose values are fixed, and the injection goes the other way, $\mathcal{K}' \hookrightarrow \mathcal{K}$.

6.2.1 Measure spaces

Experimental outcomes, whether viewed as discrete or on a continuum, are expressed in theory by points of some "outcome" set Ω equipped with a σ-algebra of measurable subsets. In the finite case of n binary detectors, any point of Ω is a string of n bits, the jth bit being 1 if the j-detector fired and otherwise 0. In this case Ω consists of the 2^n possible strings of n bits. By $\tilde{\Omega}$ we denote measurable subsets of Ω, that is, the things that are assigned probabilities by a probability measure. Whenever Ω is a finite or countably infinite set, we can take $\tilde{\Omega}$ to be all the subsets of Ω. To make our formulation more general we include uncountable outcome sets, such as the real line. When Ω is uncountable, then not all of its subsets can be measurable, so that the set $\tilde{\Omega}$ of measurable subsets of Ω is instead a σ-algebra [11].

An example of an outcome is a position as a number of meters; another example is momentum as a number of $\mathrm{kg\,m\,s^{-1}}$. We take each outcome – whether a binary number, an integer, a real number, or a list of real numbers – to be numerical; that is, we detach the outcome as a numerical entity from units, such as meters or seconds, required to interpret its significance. By this trick we avoid requiring distinct measure spaces for measurements of variables that are interpreted in distinct units.

A *probability measure* $\mu\colon \tilde{\Omega} \to [0, 1]$ is a positive measure of total measure 1; i.e., it assigns to each $\omega \in \tilde{\Omega}$ a real number between 0 and 1, subject to axioms for unions and intersections [11]. Thus by μ we denote a probability *measure*, while by $\mu(\omega)$ we denote a probability, meaning a number between 0 and 1, inclusive. We denote the set of all probability measures on measurable sets $\tilde{\Omega}$ by $\mathrm{PM}(\tilde{\Omega})$.

Describing some extensions of an experiment requires expanding the outcome set. For example, adding detectors corresponds to a new outcome set $\Omega' = \Omega \times \Omega''$, with a surjection (think of a projection) back to Ω, $\xi\colon \Omega' \to \Omega$, such that, if ω is measurable in Ω, then $\xi^{-1}(\omega)$ is measurable in Ω'; that is, we have

$$(\forall \omega \in \tilde{\Omega})\ \ \xi^{-1}(\omega) \in \tilde{\Omega}'. \tag{6.2}$$

The idea is that Ω lacks detail definable in Ω'. Here is a simple example in which Ω models a single binary detector while Ω' models two detectors: let $\Omega = \{0, 1\}$, $\Omega' = \{00, 01, 10, 11\}$, and define ξ by $\xi(ij) = i$ for $i, j \in \{0, 1\}$.

Thinking of Ω' as expressing an extension of an experiment by adding detectors, we expect a correspondence between a probability measure μ' on Ω' and μ on Ω under the surjection ξ:

$$(\forall\, \omega \in \tilde{\Omega})\ \ \mu'(\xi^{-1}(\omega)) = \mu(\omega). \tag{6.3}$$

When this holds, we say the probability measure μ' *refines* the probability measure μ via the mapping ξ.

The set of probability measures $\mathrm{PM}(\tilde{\Omega})$ can be equipped with a metric $D_{\tilde{\Omega}}$, which defines for any two measures $\mu, \lambda \in \mathrm{PM}(\tilde{\Omega})$ a distance $D(\mu, \lambda)$. An option for a metric is the L_1 distance. For this, let s denote a finite or countably infinite partition of the outcome set Ω into pairwise disjoint measurable subsets ω. Let $\boldsymbol{S}$ be the set of all such partitions. Then the L_1 distance between PMs μ and ν on $\tilde{\Omega}$ is defined by a supremum over partitions:

$$L_{1,\tilde{\Omega}}(\mu, \nu) = \frac{1}{2} \sup_{s \in \boldsymbol{S}} \sum_{\omega \in s} |\mu(\omega) - \nu(\omega)|. \tag{6.4}$$

6.2.2 Parametrized probability measures

Given a parameter space $\mathcal{K}$ and measurable sets $\tilde{\Omega}$, let a function that assigns to each $k \in \mathcal{K}$ a probability measure (PM) on $\tilde{\Omega}$ be called a *parameterized probability measure* (PPM) on the pair $(\mathcal{K}, \tilde{\Omega})$. Sometimes we look at a PPM as a function $\mathcal{K} \to \mathrm{PM}(\tilde{\Omega})$; alternatively we can see any PPM as a function $\mathcal{K} \times \tilde{\Omega} \to [0, 1]$ that assigns to k and ω a probability. We often let $\mu^{(\cdot)}$ denote a PPM; then we denote the probability measure that it assigns to the parameter list k by $\mu^{(k)}$, and we denote the probability that $\mu^{(k)}$ assigns to a measurable set ω by $\mu^{(k)}(\omega)$. In what follows it is necessary to distinguish among $\mu^{(\cdot)}$ as a PPM, $\mu^{(k)}$ as a PM, and $\mu^{(k)}(\omega)$ as a probability.

The mappings $\Xi\colon \mathcal{K} \hookrightarrow \mathcal{K}'$ and $\xi\colon \Omega' \to \Omega$ allow us to define the *envelopment* of one PPM by another. Suppose we have

$$(\forall k \in \mathcal{K})(\forall \omega \in \tilde{\Omega})\ \ \mu'^{(\Xi(k))}(\xi^{-1}(\omega)) = \mu^{(k)}(\omega). \tag{6.5}$$

In this case we say the PPM $\mu'^{(\cdot)}$ *envelops* the PPM $\mu^{(\cdot)}$ via Ξ and ξ. Entailed in this envelopment is that $\mu'^{(\Xi(k))}$ refines $\mu^{(k)}$. Except when the injection Ξ is surjective, there are elements of $\mathcal{K}' \setminus \Xi(\mathcal{K})$ that correspond to parameter values in $\mathcal{K}'$ that are unavailable in $\mathcal{K}$.

6.2.3 Quantum models

The term *model* is employed in both physics and mathematics, but differently, and within physics two uses of *model* are worth distinguishing. One use in physics

has to do with thinking about how to arrange the devices of an experiment: e.g., an experimenter may think of (model) the introduction of a polarization rotator as inserting a matrix (say as an element of the group SU(2)). In its other use in physics, which is of most concern here, "to model" is to use a PPM as a model of the relative frequencies of experimental outcomes in relation to experimentally controlled parameters. In contrast to both these uses, in mathematics, a model of a mathematical system of axioms consists of relations among entities defined in some other system of axioms.

PPMs serve as models in the sense of physics: they model relative frequencies of experimental outcomes in relation to experimentally controlled parameters. As we shall show shortly, given a probability measure, one can choose a density operation and a POVM that generate the probability measure (via (6.1)). Such a density operator and POVM constitute a model, in the sense of mathematics, of the given probability measure. By similarly dealing with probability measures, density operators, and the POVMs that are all functions of parameter lists, we arrive at *quantum models* as mathematical models of PPMs; we build up to their definition as follows.

1. By a POVM on measurable subsets $\tilde{\Omega}$ and a Hilbert space $\mathcal{H}$, we mean a function $M\colon \tilde{\Omega} \to \{\text{positive, bounded operators on } \mathcal{H}\}$, satisfying the following condition: if $\{\omega_j\}$ is a finite or countably infinite partition of the outcome set Ω into pairwise disjoint measurable subsets, then
$$\sum_j M(\omega_j) = \mathbf{1}_{\mathcal{H}}. \tag{6.6}$$
(See the discussion of "generalized observables" in [12].) We denote the set of all POVMs on $\tilde{\Omega}$ and $\mathcal{H}$ by $\mathrm{POVM}(\tilde{\Omega}, \mathcal{H})$.
2. By a density operator on $\mathcal{H}$ we mean a bounded, positive, trace-class operator having unit trace. Denote the set of all density operators on $\mathcal{H}$ by $\mathrm{DensOps}(\mathcal{H})$.
3. Suppose a PPM $\mu^{(\cdot)}$ on $(\mathcal{K}, \tilde{\Omega})$ is given. Adjusting our previous concept of a quantum model [4, 6], we now define a quantum model to be a mathematical model of a PPM, consisting of
   (a) a separable Hilbert space $\mathcal{H}_\alpha$,
   (b) a *density-operator function* $\rho_\alpha^{(\cdot)}\colon \mathcal{K} \to \mathrm{DensOps}(\mathcal{H}_\alpha)$, and
   (c) a *POVM-function* $M_\alpha^{(\cdot)}\colon \mathcal{K} \to \mathrm{POVM}(\tilde{\Omega}, \mathcal{H}_\alpha)$,
   such that
$$(\forall k \in \mathcal{K})(\forall \omega \in \tilde{\Omega})\ \ \mathrm{Tr}[\rho_\alpha^{(k)} M_\alpha^{(k)}(\omega)] = \mu^{(k)}(\omega). \tag{6.7}$$

Given a PPM $\mu^{(\cdot)}$ on $(\mathcal{K}, \tilde{\Omega})$, we define the class of models that generate this $\mu^{(\cdot)}$ to be

$$\mathrm{MODELS}(\mu^{(\cdot)}, \mathcal{K}, \tilde{\Omega}) = \{(\mathcal{H}, \rho^{(\cdot)}, M^{(\cdot)}) | (\forall \omega \in \tilde{\Omega})\ \ \mathrm{Tr}\left[\rho^{(\cdot)} M^{(\cdot)}(\omega)\right] = \mu^{(\cdot)}(\omega)\}, \tag{6.8}$$

with the understanding that, for all $k \in \mathcal{K}$, $\rho^{(k)} \in \mathrm{DensOps}(\mathcal{H})$ and $M^{(k)} \in \mathrm{POVM}(\tilde{\Omega}, \mathcal{H})$. Note that the designation $\mathrm{MODELS}(\mu^{(\cdot)}, \mathcal{K}, \tilde{\Omega})$ specifies no particular Hilbert space, and this is important: distinct models having non-isomorphic Hilbert spaces can belong to the same class of models. Usually, each parameter list k is a concatenation of a sublist k_{prep} for "state preparation" and a disjoint sublist k_{meas} for "measurement": $k = k_{\mathrm{prep}} \| k_{\mathrm{meas}}$, in which case (6.7) specializes to

$$(\forall k \in \mathcal{K})(\forall \omega \in \tilde{\Omega}) \;\; \mathrm{Tr}[\rho_\alpha^{(k_{\mathrm{prep}})} M_\alpha^{(k_{\mathrm{meas}})}(\omega)] = \mu^{(k)}(\omega). \tag{6.9}$$

6.3 Many quantum models of any PPM

When some PPM $\mu^{(\cdot)}$ on $(\mathcal{K}, \tilde{\Omega})$ has been abstracted from experimental data, one may want to choose a model to fit it. As proved in [6], this choice is highly non-unique. For instance, define the overlap of two density operators ρ and ρ' by

$$\mathrm{Overlap}(\rho, \rho') \stackrel{\mathrm{def}}{=} \mathrm{Tr}[\rho^{1/2} \rho'^{1/2}]. \tag{6.10}$$

Then, by Proposition 2 of [6], whenever $\mathrm{MODELS}(\mu^{(\cdot)}, \mathcal{K}, \tilde{\Omega})$ contains a model α with the property that for some lists k_1 and k_2 the density operators $\rho_\alpha^{(k_1)}$ and $\rho_\alpha^{(k_2)}$ have a positive overlap, $\mathrm{MODELS}(\mu^{(\cdot)}, \mathcal{K}, \tilde{\Omega})$ also contains a model β for which the overlap of $\rho_\beta^{(k_1)}$ and $\rho_\beta^{(k_2)}$ is zero. Another interesting question is this: does every PPM have a quantum model? At least when $\mathcal{K}$ is finite or countably infinite, the answer is "yes," as follows.

Proposition. Let $\mu^{(\cdot)}$ be any PPM on $(\mathcal{K}, \tilde{\Omega})$ with $\mathcal{K}$ finite or countably infinite; then $\mathrm{MODELS}(\mu^{(\cdot)}, \mathcal{K}, \tilde{\Omega})$ is not empty.

Proof. Without loss of generality, let $\mathcal{H}$ be a Hilbert space with an orthonormal basis $|j\rangle$, $j \in \mathcal{K}$. Let $\rho^{(k)} = |k\rangle\langle k|$ and define a POVM (independent of k) by

$$(\forall \omega \in \tilde{\Omega}) \;\; M(\omega) = \sum_{k' \in \mathcal{K}} \mu^{(k')}(\omega) |k'\rangle\langle k'|. \tag{6.11}$$

Then we have

$$(\forall k \in \mathcal{K})(\forall \omega \in \tilde{\Omega}) \;\; \mathrm{Tr}[\rho^{(k)} M(\omega)] = \mu^{(k)}(\omega). \tag{6.12}$$

Suppose now that $\mathcal{K} \subset \mathcal{K}_{\mathrm{prep}} \times \mathcal{K}_{\mathrm{meas}}$, so that each list of parameters splits into two sublists: $k = k_{\mathrm{prep}} \| k_{\mathrm{meas}}$. Does every PPM on such a $\mathcal{K}$ have a quantum model that respects this split, in the sense that $k_{\mathrm{prep}} \in \mathcal{K}_{\mathrm{prep}}$ specifies a density operator while $k_{\mathrm{meas}} \in \mathcal{K}_{\mathrm{meas}}$ specifies a POVM?

Proposition. Let $\mu^{(\cdot)}$ be any PPM on $(\mathcal{K}, \tilde{\Omega})$ with $\mathcal{K} = \mathcal{K}_{\mathrm{prep}} \times \mathcal{K}_{\mathrm{meas}}$; if $\mathcal{K}_{\mathrm{prep}}$ is finite or countably infinite, $\mathrm{MODELS}(\mu^{(\cdot)}, \mathcal{K}, \tilde{\Omega})$ contains a model that respects

the splitting of $\mathcal{K}$; that is, there exists a Hilbert space $\mathcal{H}$, a density-operator function $\rho^{(\cdot)}$: $\mathcal{K}_{\text{prep}} \to \text{DensOps}(\mathcal{H})$, and a POVM-function $M^{(\cdot)}$: $\mathcal{K}_{\text{meas}} \to \text{POVM}(\tilde{\Omega}, \mathcal{H})$ satisfying

$$(\forall k \in \mathcal{K})(\forall \omega \in \tilde{\Omega}) \;\; Tr[\rho^{(k_{\text{prep}})} M^{(k_{\text{meas}})}(\omega)] = \mu^{(k)}(\omega). \tag{6.13}$$

Proof. Let $\mathcal{H}$ be a Hilbert space with an orthonormal basis $|j\rangle$, $j \in \mathcal{K}_{\text{prep}}$. For any $k_{\text{prep}} \in \mathcal{K}_{\text{prep}}$, let $\rho^{(k_{\text{prep}})} = |k_{\text{prep}}\rangle\langle k_{\text{prep}}|$ and define a POVM function on $\mathcal{K}_{\text{meas}}$ by

$$(\forall \omega \in \tilde{\Omega}) \;\; M^{(k_{\text{meas}})}(\omega) = \sum_{k_{\text{prep}}} \mu^{(k_{\text{prep}} \| k_{\text{meas}})}(\omega) \left|k_{\text{prep}}\right\rangle\left\langle k_{\text{prep}}\right|. \tag{6.14}$$

(Because of the sum, $M^{(k_{\text{meas}})}$ is independent of k_{prep}.) Then we have

$$(\forall k \in \mathcal{K})(\forall \omega \in \tilde{\Omega}) \;\; \text{Tr}[\rho^{(k_{\text{prep}})} M^{(k_{\text{meas}})}(\omega)] = \mu^{(k)}(\omega). \tag{6.15}$$

Remark. If one adds the requirement that the detection operators are tensor products, as in [5], then no such proposition holds.

Remark. Belonging to the same class $\text{MODELS}(\mu^{(\cdot)}, \mathcal{K}, \tilde{\Omega})$ is a much looser relation among models than is unitary equivalence. For two models α and β of a given class, the Hilbert spaces $\mathcal{H}_\alpha$ and $\mathcal{H}_\beta$ can differ in dimension, and the inner products of pure states of model α need not match those of model β [6]. In addition, one encounters models α and β of the same class with unequal von Neumann entropies: for some $k \in \mathcal{K}$, one has

$$S(\rho_\alpha^{(k)}) \neq S(\rho_\beta^{(k)}), \tag{6.16}$$

where S denotes the von Neumann entropy of a density operator

$$S(\rho) \stackrel{\text{def}}{=} -Tr[\rho \log_2(\rho)], \tag{6.17}$$

with the understanding that $S(\rho)$ can be infinite. In particular, there are cases of two models in the same class with $S(\rho_\alpha^{(k)})$ infinite while $S(\rho_\beta^{(k)})$ is finite.

Remark. In choosing a model within a given class, the fit to probabilities offers no guidance, for all models in the class exhibit the same probability measures over $\mathcal{K}$.

Remark. Rather than choosing a density-operator function and a POVM function to fit a given PPM, often a theorist goes the other way by thinking first in terms of quantum states (for instance, wavefunctions) and operators. These can be put in the form of a pair $(\rho^{(\cdot)}, M^{(\cdot)})$ – a density-operator function and POVM function – which generate a PPM via (6.16). Indeed, this is the path of prediction. The theorist can invoke concepts of particles, leading to quantum states for electrons, photons, etc. When a PPM calculated from particle-associated states matches a subsequently

performed experiment, the PPM is confirmed as a prediction. But what does this tell us about the particles and their associated quantum states? One confirms that the particles lead to a PPM that fits the experiment, but as proved in [6], there are lots of other configurations of particles and their quantum states that serve as alternative models of the same PPM. See Ref. [13] for an example of the merit of attending to alternative hypotheses in evaluating claims for the security of quantum key distribution.

6.3.1 Application of non-uniqueness of models to the Holevo bound

Invoked in quantum communication [14] is a model α of a quantum channel, in which Alice with probability $p_\alpha^{(i)}$ prepares a state $\rho_\alpha^{(i)}$ transmitted to a receiver Bob described by a fixed POVM M_α, as illustrated in Figure 6.1.

Bob's outcome set Ω_α is finite or countably infinite and consists of disjoint measurable events $j = 1, 2, \ldots$. Model α generates the conditional probability of Bob's outcome j given that Alice prepares $\rho_\alpha^{(i)}$ as

$$\mu_\alpha^{(i)}(j) = \mathrm{Tr}[\rho_\alpha^{(i)} M_\alpha(j)]. \tag{6.18}$$

We denote the mutual information between A (for Alice) and B (for Bob) ascribed by model α by $I_\alpha(\mathrm{A}, \mathrm{B})$. The definition of this mutual information makes use of Shannon entropy, defined for any discrete probability measure μ (alternatively written $\mu(\cdot)$) by

$$H(\mu(\cdot)) \stackrel{\text{def}}{=} -\sum_j \mu(j) \log_2[\mu(j)]. \tag{6.19}$$

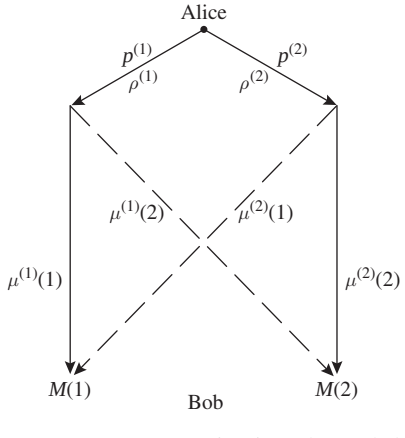


Fig. 6.1 The quantum communication channel (binary case).

Note that, if $\mu^{(i)}$ is a probability measure, so is $\sum_i p^{(i)}\mu^{(i)}$. From the definition of the mutual information [15] we have

$$I_\alpha(\mathrm{A},\mathrm{B}) = H\left(\sum_i p_\alpha^{(i)}\mu_\alpha^{(i)}(\cdot)\right) - \sum_i p_\alpha^{(i)} H\left(\mu_\alpha^{(i)}(\cdot)\right)$$
$$= H\left(\sum_i p_\alpha^{(i)} \operatorname{Tr}[\rho_\alpha^{(i)} M_\alpha(\cdot)]\right) - \sum_i p_\alpha^{(i)} H\left(\operatorname{Tr}[\rho_\alpha^{(i)} M_\alpha(\cdot)]\right). \tag{6.20}$$

For mutual information derived from density operators and POVMs, Holevo [16] showed an upper bound:

$$I(\mathrm{A},\mathrm{B}) \le \chi(\mathrm{A},\mathrm{B}) \stackrel{\text{def}}{=} S\left(\sum_i p^{(i)}\rho^{(i)}\right) - \sum_i p^{(i)} S\left(\rho^{(i)}\right), \tag{6.21}$$

where S is the von Neumann entropy defined in (6.17). This bound is useful in many cases, but, when frequency is accounted for [17], the von Neumann entropy can be arbitrarily large, indeed infinite, in which case the Holevo bound tells us nothing. Notice, however, that, in (6.20) for mutual information, the density operators enter only as factors in products with the POVM detection operators. For this reason, for any models α and β, we have

$$\mu_\alpha^{(\cdot)} = \mu_\beta^{(\cdot)} \Rightarrow I_\alpha(\mathrm{A},\mathrm{B}) = I_\beta(\mathrm{A},\mathrm{B}). \tag{6.22}$$

From this follows the tighter bound on mutual information

$$I_\alpha(\mathrm{A},\mathrm{B}) \le \min_{\beta\in[\alpha]} \chi_\beta(\mathrm{A},\mathrm{B}), \tag{6.23}$$

where we write $[\alpha]$ as shorthand for $\mathrm{MODELS}(\mu_\alpha^{(\cdot)}, \mathcal{K}, \tilde{\Omega})$.

6.3.2 An example from quantum cryptography of an enveloping PPM

The envelopment of one PPM by another has application to showing a likely vulnerability to a widely discussed design for quantum key distribution (QKD) [18]. In its simplest form, QKD calls for Alice to transmit key bits together with some extra bits to Bob. Each bit is carried by a signal modeled as a quantum state, with the promise that, if Alice and Bob compare notes on the extra bits (over a public channel), any significant eavesdropping must induce errors that Alice and Bob's comparison will reveal to them. The promise is of security against *undetected* eavesdropping.

For quantum states as light states, Bennett and Brassard [19] posit a sequence of trials, at each of which Alice prepares one of four possible single-photon light states, chosen at random. In the original design (BB84), this is expressed by a

model α with a Hilbert space $\mathcal{H}_\alpha$ taken as a two-dimensional real vector space; each bit is expressed by a density operator $\rho_\alpha^{(\theta)} = |\theta\rangle\langle\theta|$, where, for orthonormal basis states $|x\rangle$ and $|y\rangle$,

$$|\theta\rangle = \cos\theta|x\rangle + \sin\theta|y\rangle; \tag{6.24}$$

here the four allowed values of θ are given by

$$\theta \in \mathcal{K}_{\alpha,\text{prep}} = \{0, \pi/4, \pi/2, 3\pi/4\}. \tag{6.25}$$

For present purposes we consider only vulnerabilities in the face of an intercept–resend attack [18], in which an eavesdropper Eve, interposed between Alice and Bob, measures the state that Alice prepares and then sends to Bob whatever she chooses. For pointing to a vulnerability present in some implementations of BB84, it suffices to assume that (1) the transmission is lossless; (2) Eve's measurements are made by a pair of orthogonally oriented detectors of perfect efficiency and zero dark count; and (3) Eve can freely choose the orientation of her detectors, so that one detector responds optimally to light states polarized at any angle θ' of Eve's choice, while the other detector responds optimally to the light state polarized at $\theta \pm \pi/2$.

Under these assumptions, Eve's outcome set $\Omega_\alpha = \{0, 1\}$, with 1 for the event of detection for polarization angle θ' and 0 for the event of detection at a polarization angle perpendicular to θ'. Eve's measurement is expressed by the projection-valued POVM function defined on $\theta' \in \mathcal{K}_{\alpha,\text{meas}} = S^1$, where S^1 denotes the unit circle $[0 \le \theta' < 2\pi)$:

$$\begin{aligned} M_\alpha^{(\theta')}(1) &= |\theta'\rangle\langle\theta'|, \\ M_\alpha^{(\theta')}(0) &= 1 - |\theta'\rangle\langle\theta'|. \end{aligned} \tag{6.26}$$

This POVM function combines with the density-operator function to generate a PPM defined by

$$\begin{aligned} \mu_\alpha^{(\theta,\theta')}(1) &= \text{Tr}[\rho^{(\theta)} M^{(\theta')}] = |\langle\theta|\theta'\rangle|^2 = \cos[2(\theta - \theta')], \\ \mu_\alpha^{(\theta,\theta')}(0) &= 1 - \mu_\alpha^{(\theta,\theta')}(1) = 1 - |\langle\theta|\theta'\rangle|^2 = \sin[2(\theta - \theta')]. \end{aligned} \tag{6.27}$$

The claim of BB84 for security against undetected eavesdropping depends on positive inner products such as $\langle\theta = 0|\theta = \pi/4\rangle = 2^{-1/2}$. Given these inner products as they enter the PPM of (6.28), there appears no way for an eavesdropper to detect with certainty what state Alice transmits. (For the secret of how Bob manages to extract a good key, see Ref. [19].) Now comes the *implementation* of BB84 with lasers and optical fibers and so on. In this, one encounters challenges unexpressed by model α, and meeting these challenges demands models that envelop model α. Here is a case in point. BB84 calls for pulse-to-pulse changes in polarization, which are technically difficult to arrange. To evade this difficulty,

one implementation employs four separate lasers, each transmitting one of the four needed polarizations. The complication is that this implementation demands precisely matched laser frequencies, an issue invisible to model α, which leaves the frequency spectrum of the light pulses unexpressed.

To think about light frequency, we envelop $\mu_\alpha^{(\cdot)}$ by a finer-grained PPM $\mu_\beta^{(\cdot)}$ produced by a model β that explicitly expresses frequency spectra. In model β (which invokes a Hilbert space $\mathcal{H}_\beta$ of infinite dimension) a single-photon state linearly polarized at an angle θ is expressed [17] by

$$|\theta, f\rangle = \int \mathrm{d}\omega\, f(\omega) a_\theta^\dagger(\omega)|0\rangle, \tag{6.28}$$

where $a_\theta^\dagger(\omega)$ is a creation operator for light at the angular frequency ω, polarized at angle θ [17]). (Apologies for using the variable ω here for the angular frequency of a light spectrum; we depend on context to distinguish this use from our main use of ω as a measurable subset of an outcome set.) Correspondingly, we have $\rho_\beta^{(\theta,f)} = |\theta, f\rangle\langle\theta, f|$. The outcome set is the same: $\Omega_\beta = \{0, 1\}$; however, the space of preparation parameters is a lot bigger:

$$\mathcal{K}_{\beta,\mathrm{prep}} = S^1 \times \{f \,|\, f \in L^2(R), \int_I \mathrm{d}\omega\, |f(\omega)|^2 = 1\}, \tag{6.29}$$

where by $L^2(R)$ we mean square-integrable functions on the real line. Suppose we limit attention to a class of functions on R for which we have a complete set of orthonormal functions f_j. Choosing an orthonormal set of f_j to slice up the frequency band as narrowly as possible, we construct a POVM function $M_\beta^{(\theta')}$ with the same parameter space $\mathcal{K}_{\beta,\mathrm{meas}} = \mathcal{K}_{\alpha,\mathrm{meas}}$ but a larger outcome space $\Omega_\beta = \Omega_\alpha \times Z^+$ defined for $j \in Z^+ = \{1, 2, \ldots\}$ by

$$\begin{aligned} M_\beta^{(\theta')}(1, j) &= |\theta', f_j\rangle\langle\theta', f_j|, \\ M_\beta^{(\theta')}(0, j) &= |\theta'_\perp, f_j\rangle\langle\theta'_\perp, f_j|, \end{aligned} \tag{6.30}$$

where $\theta_\perp$ is an angle perpendicular to θ in the plane of linear polarization. The POVM function defined in (6.30) combines with the density-operator function to generate a PPM on $\mathcal{K}_{\beta,\mathrm{prep}} \times \mathcal{K}_{\beta,\mathrm{meas}}$, defined by

$$\begin{aligned} \mu_\beta^{(\theta,f,\theta')}(1, j) &= \mathrm{Tr}[\rho^{(\theta,f)} M^{(\theta')}(1, j)] = |\langle\theta, f|\theta', f_j\rangle|^2 \\ &= \cos^2(\theta - \theta') \left| \int_I \mathrm{d}\omega\, f^*(\omega) f_j(\omega) \right|^2, \\ \mu_\beta^{(\theta,f,\theta')}(0, j) &= |\langle\theta, f|\theta'_\perp, f_j\rangle|^2 \\ &= \sin^2(\theta - \theta') \left| \int_I \mathrm{d}\omega\, f^*(\omega) f_j(\omega) \right|^2. \end{aligned} \tag{6.31}$$

Now $\mu_\beta^{(\cdot)}$ envelops $\mu_\alpha^{(\cdot)}$ via the obvious projection of Ω_β onto Ω_α that ignores frequency spectra and, for example, the injection of $\mathcal{K}_\alpha$ into $\mathcal{K}_{\beta,\text{prep}}$ defined by $\Xi(\theta) = (\theta, f_1)$. But $\mathcal{K}_{\beta,\text{prep}}$ contains lots of elements outside the image of this injection. For example, if Alice's lasers of nominal wavelength 1.5 nm and pulses of duration 1 ns are imperfectly matched in frequency by one part in 10^{-5}, then the inner products that in model α appear to be $2^{-1/2}$ can all become essentially zero, as expressed in model β by virtue of the integrals in (6.31). Under this circumstance the transmission is vulnerable to an intercept–resend attack, because the four choices of signals for transmission, although not distinguishable clearly on the basis of polarization, become essentially perfectly distinguishable by their frequency spectra, which were ignored in model α. (For a different way in which the envelopment of one PPM by another can challenge claims of cryptographic security, see Ref. [20].)

6.4 Parameter spaces and PPMs associated with entangled states

Ten years after Born discussed them [21], interest in non-factorizable quantum states jumped when Einstein, Podolsky, and Rosen (EPR) [22] examined the theory of their measurement. Decades later interest in such states, now called *entangled*, jumped again when Bell displayed certain correlations visible in what in our words is a PPM associated with an entangled state [23] – correlations that violate inequalities satisfied by local hidden-variable theories. Later still Gisin showed that any entangled pure state can be associated with a PPM that violates generalized Bell inequalities [7, 24].

From a few well-known examples in which quantum states strongly violate Bell inequalities, we extract parameter spaces and PPMs to see patterns in places where we had never before thought to look. While a wider variety of patterns invites future attention, the patterns displayed here are primarily topological. Of interest are the topological features of the parameter spaces and also features of what we call a *level set* of the PPM over a parameter space $\mathcal{K}$, meaning a subset $[k] \subset \mathcal{K}$ over which the probability measure assigned by the PPM takes a particular value; more formally, any PPM $\mu^{(\cdot)}$ over $\mathcal{K}$ partitions $\mathcal{K}$ into level sets $[k]$ by

$$[k] \stackrel{\text{def}}{=} \{k' \in \mathcal{K} | \mu^{(k')} = \mu^{(k)}\}. \tag{6.32}$$

6.4.1 Formulation

We limit our attention to cases of a pure state $|\psi\rangle$, corresponding to density operator $\rho = |\psi\rangle\langle\psi|$, so that, for any POVM M on the Hilbert space $\mathcal{H}$ in which $|\psi\rangle$ resides, we have

$$\mathrm{Tr}[\rho M(\omega)] = \langle\psi|M(\omega)|\psi\rangle. \tag{6.33}$$

Given any tensor-product factorization of the Hilbert space $\mathcal{H}$ into

$$\mathcal{H} = \mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{B}, \tag{6.34}$$

one conventionally calls a pure state $|\psi\rangle$ *un*entangled (with respect to this factorization) if

$$(\exists|\psi_\mathrm{A}\rangle \in \mathcal{H}_\mathrm{A},\ |\psi_\mathrm{B}\rangle \in \mathcal{H}_\mathrm{B})\ \ |\psi\rangle = |\psi_\mathrm{A}\rangle \otimes |\psi_\mathrm{B}\rangle; \tag{6.35}$$

otherwise it is said to be *entangled*.

6.4.2 EPR

Reporting on the theory of measuring what we now call an entangled state, EPR speak of two particles, 1 and 2, that interact for a limited time and then separate, after which they are expressed by a wavefunction of a sort that is now called *entangled*. EPR discuss measurements made in two parts, one part at one location for particle 1, the other part at a location widely separated from the first for particle 2. Each local measurement involved a choice of parameter, corresponding in the language here of a choice of $k \in \mathcal{K}_\mathrm{meas}$. For their example of an entangled state, EPR showed that the variance of a probability measure for outcome events at one location varied with the choice of measurement made at the other location.

Following the EPR notation, we set aside for the moment our use of the letters A and B. Note also that, instead of POVMs, which came later, EPR used Hermitian operators to express measurements, so that rather than dealing with probability measures they dealt with only expectation values and variances. They wrote A for the momentum operator for particle 1, B for the position operator of particle 1, P for the momentum operator for particle 2, and Q for the position operator of particle 2. The issue is the circumstances under which the variance in the probability measure for outcomes is narrow or wide. They show that the choice of whether A or B is measured at 1 changes which of P or Q yields a narrow variance at 2.

Rephrasing this in the language of this chapter, rather than limiting attention to expectation values, variances, and possibly higher moments associated with Hermitian operators for "quantities," we note that, from each such operator, via the spectral decomposition [25], one arrives at a POVM. In this way the EPR notion of a "quantity that has with certainty the value" x translates to a probability measure having a sharp peak that concentrates its mass at or near the value x. From our viewpoint, EPR dealt with discrete parameters (k_1, k_2) with $k_1 \in \{A, B\} \cong \{0, 1\}$ and $k_2 \in \{P, Q\} \cong \{0, 1\}$. Taking the outcome sets to be $\Omega_1 = \Omega_2 = R$ (real numbers), regardless of whether the outcome pertains to meters or to momentum,

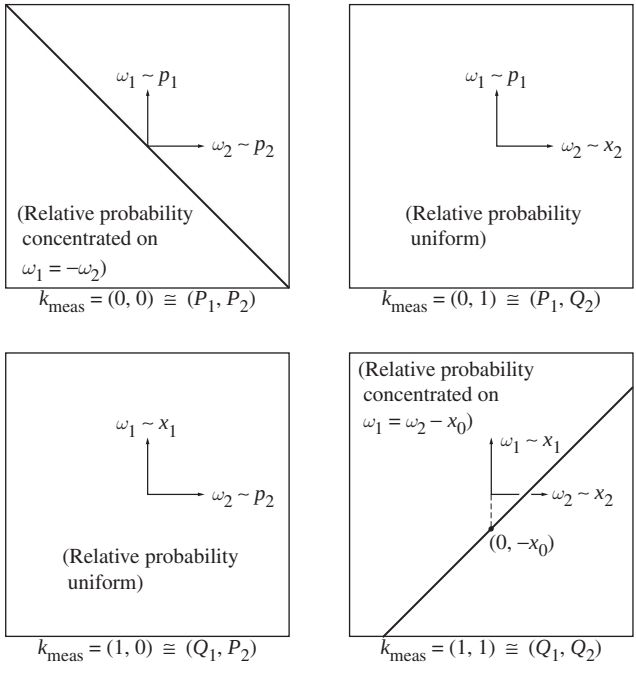


Fig. 6.2 EPR parameterized probabilities.

their state and measurement operators generate a PPM over the four combinations (k_1, k_2) as shown in Fig. 6.2. (Note that EPR used δ-functions, leading to relative probabilities; in the rest of our work, we smear any δ-functions to get plain probabilities.)

6.4.3 Generalized Bell inequalities and their violation

As with EPR, Bell's and many later studies of entangled states [24] also invoke measurement operators, mostly projections, which we view as a specialization of POVMs. Viewed this way, these studies show features of PPMs derived from separating a measurement into separate measurements in two locations A and B. The separation is expressed in part by

(1) a parameter space $\mathcal{K}_{\text{meas}} \subset \mathcal{K}_A \times \mathcal{K}_B$; and
(2) an outcome space $\Omega = \Omega_{\text{A}} \times \Omega_{\text{B}}$, each Cartesian-product factor having its own measurable subsets, $\widetilde{\Omega_{\text{A}}}$ and $\widetilde{\Omega_{\text{B}}}$, respectively.

This partitioning of parameter lists and of measurable sets implies a PPM of the form

$$\mu^{(k_{\text{A}}, k_{\text{B}})}(\omega_{\text{A}}, \omega_{\text{B}}).$$

In addition, the PPM $\mu^{(\cdot,\cdot)}$ inherits restrictions that follow from the form of any of the models involving entangled states that generate it. In any such model the separation into locations A and B finds additional expression as

(1) a Hilbert space $\mathcal{H} = \mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{B}$; and
(2) a POVM function with POVMs $M^{(k_\mathrm{A},k_\mathrm{B})} = M_\mathrm{A}^{(k_\mathrm{A})} \otimes M_\mathrm{B}^{(k_\mathrm{B})}$, where $M_\mathrm{A}^{(k_\mathrm{A})} \in \mathrm{POVM}(\widetilde{\Omega}_\mathrm{A}, \mathcal{H}_\mathrm{A})$ and $M_\mathrm{B}^{(k_\mathrm{B})} \in (\widetilde{\Omega}_\mathrm{B}, \mathcal{H}_\mathrm{B})$.

This factorization imposes on the PPM the form

$$\mu^{(k_\mathrm{A},k_\mathrm{B})}(\omega_\mathrm{A}, \omega_\mathrm{B}) = \langle\psi| M_\mathrm{A}^{(k_\mathrm{A})}(\omega_\mathrm{A}) \otimes M_\mathrm{B}^{(k_\mathrm{B})}(\omega_\mathrm{B})|\psi\rangle. \tag{6.36}$$

(Because we focus on a single state, we hold k_prep fixed and omit writing it.)

For any entangled state there exists a set of four unentangled POVMs for which the corresponding four probability measures violate a generalized Bell inequality [7]. Combined with the proof that the PPM that displays violations can be generated by disparate quantum models, this makes the PPM $\mu^{(\cdot,\cdot)}$ itself an interesting object to study.

6.4.4 The no-signaling condition

The assumption of factorization of POVMs has an implication for the probability measures involving them; namely, the well-known "no-signaling condition" is satisfied by the marginal probabilities:

$$\begin{aligned}
(\forall \omega_\mathrm{A} \in \widetilde{\Omega}_\mathrm{A})\ \ \mu^{(k_\mathrm{A},\not{k}_\mathrm{B},\ldots)}(\omega_\mathrm{A}) &\stackrel{\mathrm{def}}{=} \mu^{(k_\mathrm{A},\not{k}_\mathrm{B},\ldots)}(\omega_\mathrm{A}, \Omega_\mathrm{B}) \text{ is independent of } k_\mathrm{B}, \\
(\forall \omega_\mathrm{B} \in \widetilde{\Omega}_\mathrm{B})\ \ \mu^{(\not{k}_\mathrm{A},k_\mathrm{B},\ldots)}(\omega_\mathrm{B}) &\stackrel{\mathrm{def}}{=} \mu^{(\not{k}_\mathrm{A},k_\mathrm{B},\ldots)}(\Omega_\mathrm{A}, \omega_\mathrm{B}) \text{ is independent of } k_\mathrm{A},
\end{aligned} \tag{6.37}$$

where we put a slash through a parameter sublist to which the indicated parameterized probabilities are insensitive.

6.4.5 Case study of a toroidal parameter space

To within questions of detector inefficiencies and how to model them, a series of experiments demonstrates clear violations of Bell inequalities. We focus on one discussed in [26] and depicted in Fig. 6.3. A light source radiates entangled single-photon states in two oppositely directed beams, A and B. Each beam impinges on a measuring assembly consisting of a polarizing beam splitter followed by two sensitive light detectors, which we call 1 and 2. For purposes of appreciating the theory of entangled states, we pretend that the detectors have no dark counts and perfect efficiency and that the transmission involves no losses. Then, following [27], we arrive at a model for which the outcome set Ω_A consists of just

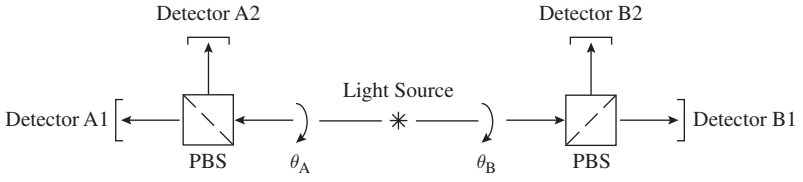


Fig. 6.3 An experiment to detect polarization-entangled photons (PBS, polarizing beam splitter).

the two points $\{1, 2\}$ that correspond to one and only one detector firing, and the same for Ω_B.

Suppose for a moment that each assembly can be rotated to any angle $\theta/2$ around the beam axis; this would correspond to detector 1 responding to light linearly polarized at angle $\theta/2$ and detector 2 responding to light linearly polarized at angle $(\theta + \pi)/2$, leading to a family of POVMs $M_A^{(\theta_A)}$ on the outcome set $\Omega_A = \{1, 2\}$, and a similar family for B. The most convenient expression of this can be found in [27], where one deals with a model light state

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|x_A\rangle|x_B\rangle + |y_A\rangle|y_B\rangle). \tag{6.38}$$

For this state one considers (projection-valued) POVMs, with the POVM $M^{(k_A)}$ (where $k_A = \theta_A$) defined by

$$\begin{aligned} M_A^{(\theta_A)}(1) &= \left(\cos\left(\frac{\theta_A}{2}\right)|x\rangle + \sin\left(\frac{\theta_A}{2}\right)|y\rangle\right)\left(\cos\left(\frac{\theta_A}{2}\right)\langle x| + \sin\left(\frac{\theta_A}{2}\right)\langle y|\right), \\ M_A^{(\theta_A)}(2) &= 1 - M_A^{(\theta_A)}(1) = \left(\sin\left(\frac{\theta_A}{2}\right)|x\rangle - \cos\left(\frac{\theta_A}{2}\right)|y\rangle\right) \\ &\quad \times \left(\sin\left(\frac{\theta_A}{2}\right)\langle x| - \cos\left(\frac{\theta_A}{2}\right)\langle y|\right). \end{aligned} \tag{6.39}$$

Replacing A by B throughout yields $M_B^{(\theta_B)}$. Via (6.1), this defines a PPM on a torus $\mathcal{K} = S^1 \times S^1$, with coordinates (θ_A, θ_B), corresponding to detectors set to respond to linearly polarized light with polarization angles $\theta_A/2$ at A and $\theta_B/2$ at B. Thus,

$$\begin{aligned} \mu^{(\theta_A,\theta_B)}(1_A, 1_B) &= \mu^{(\theta_A,\theta_B)}(2_A, 2_B) = \frac{1}{4}[1 + \cos(\theta_A - \theta_B)], \\ \mu^{(\theta_A,\theta_B)}(1_A, 2_B) &= \mu^{(\theta_A,\theta_B)}(2_A, 1_B) = \frac{1}{4}[1 - \cos(\theta_A - \theta_B)]. \end{aligned} \tag{6.40}$$

A kind of "correlation" is defined on $\mathcal{K}$ in terms of this PPM by

$$\begin{aligned} E(\theta_A, \theta_B) &= \mu^{(\theta_A,\theta_B)}(1_A, 1_B) + \mu^{(\theta_A,\theta_B)}(2_A, 2_B) - \mu^{(\theta_A,\theta_B)}(1_A, 2_B) - \mu^{(\theta_A,\theta_B)}(2_A, 1_B) \\ &= \cos(\theta_A - \theta_B). \end{aligned} \tag{6.41}$$

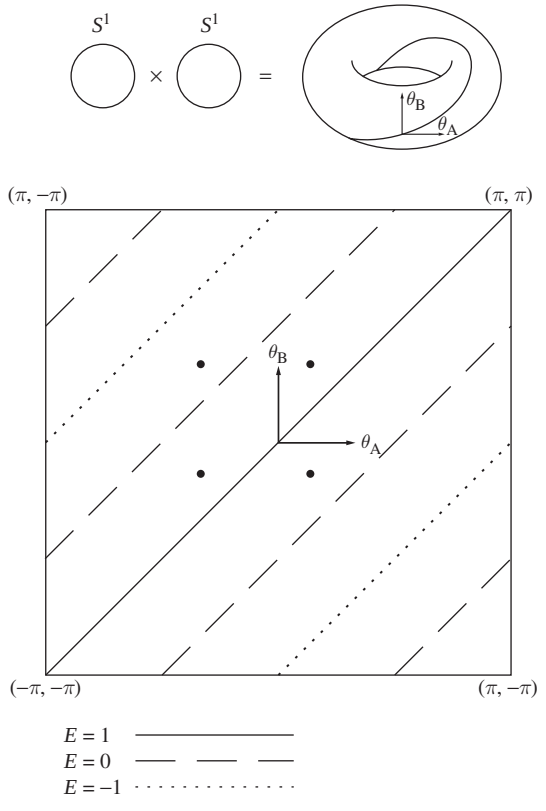


Fig. 6.4 Parameter space $S^1 \times S^1$ and contours of constant probability measure E on a torus with marked points $(\theta_A, \theta_B) = \{(-3\pi/8, -\pi/8), (\pi/8, -\pi/8), (\pi/8, 3\pi/8), (-3\pi/8, 3\pi/8)\}$ for a maximal violation of the Bell inequality.

The generalized Bell inequality at issue involves the correlation E evaluated at four pairs of values of the angles:

$$-2 \leq S_{\text{Bell}} \leq 2, \tag{6.42}$$

for

$$S_{\text{Bell}} \stackrel{\text{def}}{=} E(\theta_A, \theta_B) - E(\theta_A, \theta'_B) + E(\theta'_A, \theta_B) + E(\theta'_A, \theta'_B). \tag{6.43}$$

As shown at the top in Fig. 6.4, the parameter space with points (θ_A, θ_B) is a torus. By skinning the surface of the donut and stretching it out, we get the square area in Fig. 6.4, which shows the contours of constant μ, namely the lines $\theta_A - \theta_B =$ constant. The level sets of $\mu^{(\cdot)}$ as a function on $\mathcal{K}$ are thus closed loops that wind once around the torus. Figure 6.4 also shows values of θ_A and θ_B for which S_{Bell} maximally violates the inequality.

6.4.6 Allowing for elliptical polarization: SO(3)

Linear polarization is needlessly restrictive. To display the drama ensuing from general elliptical polarization, we need a different light state. By inserting a suitable element in the path toward detector B, one can change the light modeled by the $|\psi\rangle$ defined in (6.38) into something modeled by a singlet state

$$|\psi'\rangle = \frac{1}{\sqrt{2}}(|x_{\rm A}\rangle|y_{\rm B}\rangle - |y_{\rm A}\rangle|x_{\rm B}\rangle), \tag{6.44}$$

where now we view $|x\rangle$ and $|y\rangle$ as orthonormal vectors in the complex vector space C^2. The advantage of the state $|\psi\rangle$ is its invariance under the same SU(2) transformation applied to both the A- and the B-factor states. We now expand the parameter space $\mathcal{K}_{\rm meas}$ to allow for detections not just of linearly polarized light but also of light of arbitrary elliptical polarization. With the same assumptions as above about perfectly efficient detectors and no dark counts, one has a detector assembly modeled by a POVM with an additional parameter ϕ:

$$\begin{aligned} M_{\rm A}^{(\theta_{\rm A},\phi_{\rm A})}(1) &= \left(\cos\left(\frac{\theta_{\rm A}}{2}\right)|x\rangle + {\rm e}^{{\rm i}\phi_{\rm A}}\sin\left(\frac{\theta_{\rm A}}{2}\right)|y\rangle\right) \\ &\quad\times\left(\cos\left(\frac{\theta_{\rm A}}{2}\right)\langle x| + {\rm e}^{-{\rm i}\phi_{\rm A}}\sin\left(\frac{\theta_{\rm A}}{2}\right)\langle y|\right), \end{aligned}$$

$$\begin{aligned} M_{\rm A}^{(\theta_{\rm A},\phi_{\rm A})}(2) &= 1 - M_{\rm A}^{(\theta_{\rm A},\phi_{\rm A})}(1) \\ &= \left(\sin\left(\frac{\theta_{\rm A}}{2}\right)|x\rangle - {\rm e}^{{\rm i}\phi_{\rm A}}\cos\left(\frac{\theta_{\rm A}}{2}\right)|y\rangle\right) \\ &\quad\times\left(\sin\left(\frac{\theta_{\rm A}}{2}\right)\langle x| - {\rm e}^{-{\rm i}\phi_{\rm A}}\cos\left(\frac{\theta_{\rm A}}{2}\right)\langle y|\right), \end{aligned} \tag{6.45}$$

together with the same expression in which A is replaced by B throughout. Here the ranges of the parameters are $0 \le \theta_{\rm A,B} \le \pi$ and $0 \le \phi < 2\pi$.

The general unit ray for an elliptically polarized state, which is represented by $\cos(\theta_{\rm A}/2)|x\rangle + {\rm e}^{{\rm i}\phi_{\rm A}}\sin(\theta_{\rm A}/2)|y\rangle$, corresponds to a point on the Poincaré sphere: ϕ corresponds to longitude while θ corresponds to latitude measured down from the North pole. By this mapping, the orthogonal state goes to the polar opposite point on the sphere. The selection of a POVM thus corresponds to the selection of two points, with coordinates $(\theta_{\rm A}, \phi_{\rm A})$ on a sphere for A, and $(\theta_{\rm B}, \phi_{\rm B})$ on a sphere for B. Instead of the torus $S^1 \times S^1$ as the parameter space when polarization is constrained to be linear, for elliptical polarization we find a parameter space that is a product of two spheres:

$$\mathcal{K}_{\rm meas} = S^2 \times S^2. \tag{6.46}$$

It is straightforward to calculate the PPM on $S^2 \times S^2$ for this case:

$$\begin{aligned}
\mu^{(\theta_A,\phi_A,\theta_B,\phi_B)}(1_A, 1_B) &= \mu^{(\theta_A,\phi_A,\theta_B,\phi_B)}(2_A, 2_B) \\
&= \frac{1}{4}[1 - \cos\theta_A \cos\theta_B - \sin\theta_A \sin\theta_B \cos(\phi_A - \phi_B)], \\
\mu^{(\theta_A,\phi_A,\theta_B,\phi_B)}(1_A, 2_B) &= \mu^{(\theta_A,\phi_A,\theta_B,\phi_B)}(2_A, 1_B) \\
&= \frac{1}{4}[1 + \cos\theta_A \cos\theta_B + \sin\theta_A \sin\theta_B \cos(\phi_A - \phi_B)].
\end{aligned} \tag{6.47}$$

On recognizing a little spherical trigonometry, one notices the following about the level sets of probability measures of the PPM defined by (6.47) on $\mathcal{K}_{\text{meas}} = S^2 \times S^2$. We can view $S^2 \times S^2$ as a space in which each element is a pair of points, (θ_A, ϕ_A) and (θ_B, ϕ_B), on one unit sphere. Let ζ be the angle between the unit vectors to these points; $0 \leq \zeta \leq \pi$. Then we have

$$\cos\zeta = \cos\theta_A \cos\theta_B + \sin\theta_A \sin\theta_B \cos(\phi_A - \phi_B), \tag{6.48}$$

so that (6.47) simplifies to

$$\begin{aligned}
\mu^{(\theta_A,\phi_A,\theta_B,\phi_B)}(1_A, 1_B) = \mu^{(\theta_A,\phi_A,\theta_B,\phi_B)}(2_A, 2_B) &= \frac{1}{4}(1 - \cos\zeta), \\
\mu^{(\theta_A,\phi_A,\theta_B,\phi_B)}(1_A, 2_B) = \mu^{(\theta_A,\phi_A,\theta_B,\phi_B)}(2_A, 1_B) &= \frac{1}{4}(1 + \cos\zeta).
\end{aligned} \tag{6.49}$$

This PPM maximally violates a Bell inequality.

In the previous case, limited to linear polarization, all the level sets of probability measures on the torus $S^1 \times S^1$ were, apart from location on the torus, the same one-dimensional loop; but here with the parameter space $S^2 \times S^2$ we have some variety. Any pair of points defines an angle ζ between the rays from the origin of the sphere to the points, and any level set for probability measures over $\mathcal{K} = S^2 \times S^2$ is the locus of all pairs having any fixed angle ζ between them, $0 \leq \zeta \leq \pi$. For the open interval $0 < \zeta < \pi$, this level set is a manifold of dimension 3 obtained as follows. Starting from any pair of points (θ_A, ϕ_A) and (θ_B, ϕ_B) on the unit sphere separated by angle ζ in the open interval, every other pair separated by ζ can be reached by one and only one rotation acting on the starting pair, so the level set is the orbit under an effective action of the rotation group SO(3). The manifold of all these pairs – the level set – is isomorphic to SO(3) as a Lie group, which, like the circle, is connected but not simply connected. In contrast for $\zeta = 0$ (where $(\theta_A, \phi_A) = (\theta_B, \phi_B)$) or for $\zeta = \pi$ (where they are polar images of one another), SO(3) has an isotropy group of SO(2) and instead of the three-dimensional manifold of SO(3), the level set is just the simply connected two-dimensional manifold, the sphere S^2.

6.4.7 The property of local reach

The PPMs of the preceding two subsections dealing with violations of Bell inequalities have a special property (not shared by some other PPMs derived from entangled states): by holding k_A fixed at any one value and varying k_B (or vice versa) one covers the whole space of measures. For example, the probability measures involved in the simplest model above of polarization-entangled light are functions of $(k_A - k_B)$, where $k_A \equiv \theta_A$ and $k_B \equiv \theta_B$ are single numbers that express the angles to which polarizing filters are set [3]. More generally we will say that a PPM $\mu^{(\cdot,\cdot)}$ has the property of "local reach" if

$$(\forall k_A, k_B, k'_B)(\exists\, k'_A)\ \ \mu^{(k_A,k_B)} = \mu^{(k'_A,k'_B)}, \tag{6.50}$$

$$(\forall k_A, k_B, k'_A)(\exists\, k'_B)\ \ \mu^{(k_A,k_B)} = \mu^{(k'_A,k'_B)}. \tag{6.51}$$

Proposition. Consider any PPM $\mu^{(\cdot)}$ on some $\mathcal{K} = \mathcal{K}_{\text{prep}} \times \mathcal{K}_{\text{meas}}$, with $\mathcal{K}_{\text{meas}} = \mathcal{K}_A \times \mathcal{K}_B$, and suppose that this PPM satisfies the no-signaling condition and has the property of local reach; then the marginal probabilities for all its probability measures are invariant over $\mathcal{K}_{\text{meas}}$.

Proof. Evaluate (6.51) for the probability of (ω_A, Ω_B) to get

$$(\forall k_A, k_B, k'_A)(\exists k'_B)(\forall \omega_A \in \widetilde{\Omega_A})\ \ \mu^{(k_A,k_B)}(\omega_A, \Omega_B) = \mu^{(k'_A,k'_B)}(\omega_A, \Omega_B). \tag{6.52}$$

By virtue of the assumed no-signaling equation, the marginal probability is insensitive to k_B, whence we have

$$(\forall k_A, k'_A)(\forall \omega_A \in \widetilde{\Omega_A})\ \ \mu^{(k_A,\not{k}_B)}(\omega_A, \Omega_B) = \mu^{(k'_A,\not{k}_B)}(\omega_A, \Omega_B), \tag{6.53}$$

where the slash through k_B indicates that its value makes no difference. Hence we have that $\mu^{(k_A,\not{k}_B)}(\omega_A, \Omega_B)$ depends neither on k_A nor on k_B; thus we can write $\mu^{(\not{k}_A,\not{k}_B)}(\omega_A, \Omega_B)$. The same argument holds for $\mu^{(\not{k}_A,\not{k}_B)}(\Omega_A, \omega_B)$.

6.4.8 Allowing for light frequency, etc.

As discussed in the example of quantum cryptography in Section 6.3.2, light involves more degrees of freedom than just polarization. Not only do light pulses come with frequency spectra, implying an infinite-dimensional Hilbert space, but also models can posit states of more than one photon [17]. As we saw in Section 6.3.2, under various special assumptions, such as that of frequency-independent detectors, models invoking the infinite-dimensional spaces provide PPMs that envelop those discussed in the preceding two subsections. These finer-grained models provide a much bigger class of PPMs, the characterization of which awaits future work.

6.4.9 Metrical properties

Bell inequalities express metrical properties of $\mu^{(\cdot)}$. For instance, as it works in the example of a toroidal parameter space, (6.42) expresses an L_1 distance between $\mu^{(\theta_A,\theta_B)}$ and $\mu^{(\theta_A,\theta'_B)}$ together with more complicated properties that quantify how much and in what ways a PPM varies both with k_{meas} and with ω. It is worth noting that the hidden-variable model discussed in [27] has the same toroidal parameter space as discussed in Section 6.4.5, but with contours such that no violation of the Bell inequality (6.42) is possible.

As is well known in quantum decision theory [14], a metric on $\text{PM}(\tilde{\Omega})$ also provides a basis for deciding between one value of a parameter and another value. For example, consider the situation a little more general than that sketched in Figure 12.1. This situation, which arises not only when Alice and Bob are communicating parties, but also when they are opponents, is described by a PPM $\mu^{(\cdot)}$ over some $(\mathcal{K}, \widetilde{\Omega_B})$ with $\mathcal{K} = \mathcal{K}_A \times \mathcal{K}_B$. Alice chooses $k_A \in \mathcal{K}_A$, Bob chooses $k_B \in \mathcal{K}_B$. Then Bob's probability of outcome $\omega_B \in \widetilde{\Omega_B}$ is $\mu^{(k_A,k_B)}(\omega_B)$. How well can Bob distinguish Alice's parameter values k_A and k'_A? We say that these values are ϵ-separable for Bob's choice of measurement k_B when

$$\epsilon \leq D_{\tilde{\Omega}}\Big(\mu^{(k_A,k_B)}, \mu^{(k'_A,k_B)}\Big). \tag{6.54}$$

As an avenue for future work, we are interested in the case in which the preparation parameters to be decided include space-time coordinates.

Also for future exploration is the issue of PPMs that are "close together." To quantify the difference between two such PPMs over a given parameter set and given measurable sets, one defines a metric on them. A plausible metric is based on the value of k at which the two probability measures most differ. That is, given a distance $D_{\tilde{\Omega}}$ on probability measures, we get a distance between two PPMs. For $\mu^{(\cdot)}$ and $\mu'^{(\cdot)}$ over $\mathcal{K} \times \tilde{\Omega}$ a distance between them can be defined by

$$\mathcal{D}\left(\mu^{(\cdot)}, \mu'^{(\cdot)}\right) \stackrel{\text{def}}{=} \sup_k D_{\tilde{\Omega}}\left(\mu^{(k)}, \mu'^{(k)}\right). \tag{6.55}$$

Such a distance is useful in comparing a PPM from a model against a second PPM extracted from experimental relative frequencies.

Acknowledgments

We thank Tai Tsun Wu, Howard Brandt, and Samuel Lomonaco for vigorous discussions of the linking of equations to experiments. JM thanks Horace Yuen for the chance to talk about an earlier version of this paper during a visit to Northwestern University. He also thanks Prem Kumar for helpful discussions, David Mumford

for an illuminating discussion of $S^3 \times S^3$, and Martin Jaspan for the invitation to participate in experiments on quantum key distribution.

References

[1] P. A. M. Dirac, *The Principles of Quantum Mechanics*, 4th edn. (Oxford: Clarendon Press, 1958).
[2] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Berlin: Springer, 1932); translated with revisions by the author as *Mathematical Foundations of Quantum Mechanics* (Princeton, NJ: Princeton University Press, 1955).
[3] A. Peres, *Quantum Theory: Concepts and Methods* (Dordrecht: Kluwer, 1993).
[4] J. M. Myers and F. H. Madjid, A proof that measured data and equations of quantum mechanics can be linked only by guesswork, in *Quantum Computation and Information*, S. J. Lomonaco Jr. and H. E. Brandt (eds.) (Providence, RI: American Mathematical Society, 2002), pp. 221–224.
[5] B. S. Tsirelson (Cirel'son) Quantum generalizations of Bell's inequality, *Lett. Math. Phys.* **4**, 93–100 (1980).
[6] F. H. Madjid and J. M. Myers, Matched detectors as definers of force, *Ann. Phys. (NY)* **319**, 251–273 (2005).
[7] N. Gisin, Bell's inequality holds for all non-product states, *Phys. Lett. A* **154**, 201–202 (1991). (Note that this title fails to say what Gisin means: where he says *holds* he means *is violated*; also his quantification of a violation has typographical errors and should be corrected to read $|P(a,b) - P(a,b')| + P(a',b) + P(a',b') = 2\left(1 + 4|c_1 c_2|^2\right)^{1/2}$.)
[8] A. Shimony, Degree of entanglement, *Ann. N.Y. Acad. Sci.* **755**, 675–679 (1995).
[9] S. Popescu and D. Rohrlich, Generic quantum non-locality, *Phys. Lett. A* **166**, 293–297 (1992).
[10] M. V. Berry, Quantal phase factors accompanying adiabatic changes, *Proc. Roy Soc. A* **392**, 45–57 (1984).
[11] W. Rudin, *Real and Complex Analysis*, 3rd edn. (New York: McGraw-Hill, 1987).
[12] A. S. Holevo, *Statistical Structure of Quantum Theory* (Berlin: Springer, 2001).
[13] J. M. Myers, Polarization-entangled light for quantum key distribution: how frequency spectrum and energy affect statistics, in *Quantum Information and Computation III*, E. J. Donkor, A. R. Pirich, H. E. Brandt (eds.) (Bellingham, WA: SPIE 2005), pp. 13–26.
[14] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press, 2000).
[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (New York: Wiley, 1991).
[16] A. S. Holevo, Some estimates for the information content transmitted by a quantum communication channel, *Probl. Pered. Inform.* **9**, 3–11 (1973). (English translation *Probl. Inf. Transm. (USSR)* **9**, 177–183 (1973).)
[17] J. M. Myers, Framework for quantum modeling of fiber-optical networks, Parts I and II, arXiv:quant-ph/0411107 v2 and quant-ph/0411108 v2 (2005).
[18] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Quantum cryptography, *Rev. Mod. Phys.* **74**, 145–195 (2002).
[19] C. H. Bennett and G. Brassard, Quantum cryptography: public key-distribution and coin tossing, in *Proceedings of the IEEE International Conference on Computers,*

Systems and Signal Processing, Bangalore, India (New York: IEEE, 1984), pp. 175–179.
[20] A. Acín, N. Gisin, and L. Masanes, From Bell's theorem to secure quantum key distribution, arXiv:quant-ph/0510094 (2005).
[21] M. Born, Zur Quantenmechanik der Stoßvorgänge, *Z. Phys.* **37**, 863–867 (1926). English translation as On the quantum mechanics of collisions, in *Quantum Theory and Measurement*, J. A. Wheeler and W. H. Zurek (eds.) (Princeton, NJ: Princeton University Press, 1983).
[22] A. Einstein, B. Podolsky, and N. Rosen, Can quantum mechanical description of physical reality be considered complete?, *Phys. Rev.* **47**, 777–780 (1935).
[23] J. S. Bell, On the Einstein–Podolsky–Rosen paradox, *Physics* **1**, 195–200 (1964).
[24] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Proposed experiment to test local hidden-variable theories, *Phys. Rev. Lett.* **23**, 880–884 (1969).
[25] W. Rudin, *Functional Analysis*, 2nd edn. (New York: McGraw-Hill, 1991), p. 368.
[26] A. Aspect, P. Grangier, and G. Roger, Experimental realization of Einstein–Podolsky–Rosen *Gedankenexperiment*, *Phys. Rev. Lett.* **49**, 91–94 (1982).
[27] A. Aspect, Experimental tests of Bell's inequalities, in *Atomic Physics 8*, I. Lindgren, A. Rosén, and S. Svanberg (eds.) (New York: Plenum Press, 1983), pp. 103–127.

7

Bayesian updating and information gain in quantum measurements

Leah Henderson

7.1 Introduction

In many situations, learning from the results of measurements can be regarded as updating one's probability distributions over certain variables. According to Bayesians, this updating should be carried out according to the rule of conditionalization. In the theory of quantum mechanics, there is a rule that tells us how to update the *state* of a system, given observation of a measurement result. The state of a quantum system is closely related to probability distributions over potential measurements. Therefore we might expect there to be some relation between Bayesian conditionalization and the quantum state-update rule. There have been several suggestions that the state change just is Bayesian conditionalization, appropriately understood, or that it is closely analogous.

Bub was the first[1] to make the connection between quantum measurement and Bayesian conditionalization in a 1977 paper ([1], see also [2]), using an approach based on quantum logic. The connection is renewed in discussions by Fuchs [3] and also Jacobs [4] in 2002, where again the analogy between the quantum state update and Bayesian conditionalization is pointed out. At the same time, Fuchs draws attention to a disanalogy – namely that there is an "extra unitary" transformation as part of the measurement in the quantum case. In this chapter, I will first review the proposals of Bub, Jacobs, and Fuchs. I will then show that the presence of the extra unitaries in quantum measurement leads to a difference between classical and quantum measurement in terms of information gain, drawing on results by Nielsen [5] and Fuchs and Jacobs [6].

We begin with a brief introduction to Bayesian conditionalization in Section 7.2, and quantum measurement in Section 7.3.

[1] To the best of my knowledge.

Philosophy of Quantum Information and Entanglement, ed. A. Bokulich and G. Jaeger. Published by Cambridge University Press.

7.2 Bayesian conditionalization

Suppose we have an initial probability distribution $p(h)$ over some propositions h in a hypothesis space H. If we gather new evidence, given by proposition d, then the initial probability distribution may need to be updated. The rule of Bayesian conditionalization says that the initial probability distribution $p(h)$ (known as the "prior") should be replaced by the conditional distribution $p(h|d)$, (called the "posterior"),

$$p(h) \rightarrow p(h|d). \tag{7.1}$$

The posterior distribution $p(h|d)$ can be calculated using Bayes' rule

$$p(h|d) = \frac{p(d|h)p(h)}{p(d)}, \tag{7.2}$$

where $p(h)$ is the prior probability distribution over the hypotheses, $p(d|h)$ is known as the likelihood of the evidence d, and $p(h|d)$ is the posterior probability. The denominator $p(d)$ is given by

$$p(d) = \sum_h p(d|h)p(h). \tag{7.3}$$

The posterior probability for h will be higher if there is a high prior probability for h or if the hypothesis renders the observed evidence d highly likely (the likelihood $p(d|h)$ is high). Also, if the evidence d is initially rather improbable (low $p(d)$), then it is relatively unlikely to be obtained given any of the alternative hypotheses. Therefore in this case too, the posterior probability will be higher.

The most common interpretation of the probabilities here is that they represent degrees of belief in the proposition held by some agent, perhaps idealized. Then Bayesian conditionalization can be taken to characterize the process by which an agent learns in a rational manner from measurement results given by d.

7.3 Quantum measurement

In quantum mechanics, the state of a system is updated after a measurement has been performed on it. To specify a measurement of a particular physical quantity, one first needs to know the possible measurement outcomes and their probabilities. The most general quantum measurements can be described using the formalism of positive operator-valued measures (POVMs). A measurement $\mathcal{A}$ is represented by positive operators E_d, which satisfy the completeness equation $\sum_d E_d = 1$. Possible outcomes of the measurement are denoted by d, and associated with each operator E_d.

There is a quantum-mechanical rule for how the state of the system changes on measurement. In the POVM formalism, if outcome d is obtained, then the system after measurement is described by a new density matrix ρ'_d,

$$\rho'_d = \frac{1}{p_d} \sum_{d'} A_{dd'} \rho A_{dd'}^\dagger, \tag{7.4}$$

where $A_{dd'}$ are measurement operators such that

$$E_d = \sum_{d'} A_{dd'}^\dagger A_{dd'}. \tag{7.5}$$

An important special case of POVMs is projective measurements. Such measurements are represented by positive operators that are orthogonal projectors, P_d. The projectors satisfy the orthogonality relation $P_d P_{d'} = \delta_{dd'} P_d$. After obtaining the result d, the state of the system is projected into the state associated with the projector P_d.

It will be useful later to distinguish between "efficient" and "inefficient" measurements. In an efficient measurement, no information is lost. If the initial state is pure, the state of the system after measuring any result will still be pure after the measurement. Projective measurements are all efficient. Efficient measurements can be expressed using just one index, so that the post-measurement states are

$$\rho'_d = \frac{A_d \rho A_d^\dagger}{p_d}. \tag{7.6}$$

In an inefficient measurement, on the other hand, the final state is a sum of different possible states indexed by d'. By not distinguishing these states, an inefficient measurement throws away some of the information that would in principle have been available.

7.4 Quantum measurement as Bayesian updating

We now review the proposals by Bub [1, 2] and Fuchs [3] for connecting the quantum state-update rule with Bayesian conditionalization. As we have seen, the usual view of Bayesian probabilities is that they are subjective degrees of belief. Both Bub and Fuchs take this view of the probabilities associated with a quantum state.

In Bub's view, quantum-mechanical probabilities are different from classical probabilities because the algebra of events on which they are defined is different. In the usual probability model, due to Kolmogorov, propositions are represented by subsets of a space, and probabilities are defined on these subsets, which form a Boolean algebra. However, the algebra of quantum-mechanical propositions is non-Boolean. One approach then is to generalize the usual probability

model so that it is defined on a non-Boolean algebra. Then quantum probabilities become generalized probabilities. Bub suggests that, when the appropriate connection between the quantum state and quantum probabilities is made, quantum state updating does correspond exactly to Bayesian conditionalization of the generalized probabilities on the non-Boolean quantum algebra.

We first briefly review some of the background from quantum logic and generalized probability, and then outline Bub's strategy. This involves defining a classical operator playing the role of the quantum density operator, and considering how it is updated when probabilities are conditionalized. The update rule for the quantum density operator is then argued to be the natural analog of the update rule for the classical operator.

Both classically and in quantum mechanics, probabilities may be assigned to propositions such as "the value of the physical quantity A lies in the range Δ." Such propositions can be combined using the standard logical connectives $\wedge$ ("and"), and $\vee$ ("or").

Classically, the proposition can be represented as the subset of those states in the state space which are such that, if A were measured on the system in the state, the result would be within the range Δ. In this representation, the logical connective $\wedge$ ("and") corresponds to set intersection $\cap$, and $\vee$ ("or") corresponds to set union $\cup$. The propositions or subsets form a Boolean algebra.

The standard model for probability is the Kolmogorov probability space. Suppose E is a set of elementary events. A field on E is a set of subsets of E that has E as a member, and that is closed under complementation and finite union. A Kolmogorov probability space is a triple $\langle E, \mathcal{F}, p\rangle$, where E is a set, $\mathcal{F}$ is a field of subsets of E, and p is a mapping $p\colon \mathcal{F} \to [0, 1]$ satisfying the properties

- $p(E) = 1$
- $p(A_i) \geq 0$ for all $A_i \in \mathcal{F}$
- $p(A_i \cup A_j) = p(A_i) + p(A_j)$ for all $A_i, A_j \in \mathcal{F}$ such that $A_i \cap A_j = \varnothing$.

The mapping p is known as a "probability function." It is common to extend these axioms to the case in which E is infinite by taking $\mathcal{F}$ to be a σ-field and extending the final axiom to countable addivitity.

- If $\{A_i\}$ is any countable family of pairwise disjoint members of $\mathcal{F}$ that are disjoint subsets, ($A_i \cap A_j = \varnothing$ whenever $i \neq j$) then $p(\cup_i \{A_i\}) = \sum_i p(A_i)$.

In quantum mechanics, the proposition "the value of the physical quantity A lies with in the range Δ" is not represented by a subset of the state space, as it is classically, but rather by a subspace of the Hilbert space, $\mathcal{H}$. In this representation, the logical connective $\wedge$ ("and") corresponds to intersection of subspaces, and $\vee$ ("or") corresponds to the span of subspaces. Let us call the set of subspaces of $\mathcal{H}$, $L(\mathcal{H})$.

The algebraic structure of $L(\mathcal{H})$ is non-Boolean, unlike the algebraic structure of subsets in the classical case. Notably, it violates the distributivity property

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c), \tag{7.7}$$

which is an axiom of a Boolean algebra. For a fuller explanation of the difference between the algebras of quantum and classical events, see [7], or [8].

When an algebra is Boolean, it is isomorphic to a field of subsets,[2] which is what classical Kolmogorov probabilities must be defined over. When the algebra is non-Boolean, there is no longer necessarily such an isomorphism. It is possible, however, to define a generalized probability model over $L(\mathcal{H})$. A generalized probability function on $L(\mathcal{H})$ is a mapping $p :\rightarrow [0, 1]$ satisfying the properties

- $p(I) = 1$, where I is the identity mapping on $\mathcal{H}$
- if $P_i \in L(\mathcal{H})$ are orthogonal projectors, then

$$p(\oplus_i P_i) = \sum_i p(P_i). \tag{7.8}$$

Suppose now that we want to represent the state of a system, from which we can generate the probabilities for the outcome of any measurement. The state in quantum mechanics can be described by a density operator ρ that is a weighted combination of the one-dimensional projectors spanning the space, which each correspond to an elementary event

$$\rho = \sum_j w_j P_j. \tag{7.9}$$

Classically, a similar operator can also be defined, where the P_j are replaced by indicator functions I_j for the sets associated with the elementary classical events X_j:

$$I_j(x) = \begin{cases} 1 & \text{if } x \in X_j, \\ 0 & \text{otherwise.} \end{cases} \tag{7.10}$$

Then the classical operator ρ is

$$\rho = \sum_j w_j I_j, \tag{7.11}$$

where $\sum_j w_j = 1$, $w_j \geq 0$, $\forall j$. Therefore $\rho(x_k) = \sum_j w_j I_j(x_k)$.

In both the classical and quantum cases, the probabilities for outcomes of a measurement on a system in the state can be generated from this operator. In quantum mechanics, as we have seen, the measurement $\mathcal{A}$ is represented by a set of POVM

[2] Stone's theorem, see [7], p. 186.

operators $\{E_d\}$. The probability of obtaining outcome d on measuring $\mathcal{A}$ is given by the trace rule:

$$p_d = \mathrm{tr}(\rho E_d), \tag{7.12}$$

where ρ is the density operator representing the initial state of the system.

Classically, the probabilities for the outcomes of a measurement can be generated from the classical operator ρ. Suppose we have a set of mutually exclusive and collectively exhaustive events that comprise the possible outcomes of a measurement. We denote these as $a_1, a_2, \ldots$ with indicator functions I_{a_i} satisfying $\sum_i I_{a_i} = 1$ and $I_{a_i} I_{a_j} = 0$ for $i \neq j$. The classical operator ρ generates a probability for these events according to

$$p_\mu(a) = \mu(X_a) = \sum_j \rho(x_j) I_a(x_j) \tag{7.13}$$

or

$$p_\rho(a) = \sum \rho I_a \tag{7.14}$$

for short.[3]

We now determine the update in the classical operator ρ which is associated with Bayesian updating on the probabilities. The conditional probability of an outcome b given a measurement of a_i is

$$p_\rho(b|a_i) = \frac{\sum \rho I_{a_i} I_b}{\sum \rho I_{a_i}}. \tag{7.15}$$

Bayesian updating

$$p_\rho(b) \to p_\rho(b|a_i) = \sum \rho' I_b \tag{7.16}$$

corresponds to a transition in the classical operator

$$\rho \to \rho' = \frac{\rho I_{a_i}}{\sum \rho I_{a_i}}. \tag{7.17}$$

Bub then claims that the symmetric expression

$$\rho \to \rho' = \frac{I_{a_i} \rho I_{a_i}}{\sum I_{a_i} \rho I_{a_i}} \tag{7.18}$$

is the update rule which is appropriate to represent conditionalization with respect to a_i when the algebra may be non-commutative. Therefore, he says, the quantum state-update rule

$$\rho \to \rho' = \frac{P_{a_i} \rho P_{a_i}}{\mathrm{tr}(P_{a_i} \rho P_{a_i})} \tag{7.19}$$

[3] See [2], p. 85.

that specifies the change in the density operator ρ on measurement of A with result a_i is just the conditionalization of ρ with respect to event a_i ([2], p. 87).

Bub's account applies only to projective measurements. It is interesting to consider the classical analog of the POVMs. In a POVM, the subspaces corresponding to the different measurement outcomes need not be orthogonal. This corresponds to a classical measurement where the outcomes of the measurement are not necessarily mutually exclusive. This is expressed by the condition on indicator functions that $I_{a_i} I_{a_j} = 0$ for $i \neq j$. We now drop this restriction, while still maintaining $\sum_i I_{a_i} = 1$. A function that satisfies these constraints is a conditional probability $p(a_i|x_j)$, where we no longer restrict the possible values to zero and one. This means that an elementary event is no longer definitely in the subset corresponding to a measurement outcome – rather there is some probability that it is in this subset, rather than in that corresponding to a different measurement outcome. On replacing $I_{a_i}(x)$ by $p(a_i|x_j)$, (7.13) becomes

$$p_\rho(a) = \sum_j \rho(x_j) p(a|x_j). \tag{7.20}$$

The conditional probability of an outcome b given a measurement of a_i is

$$p_\rho(b|a_i) = \frac{\sum_j \rho(x_j) p(a_i|x_j) p(b|x_j)}{\sum_j \rho(x_j) p(a_i|x_j)} \tag{7.21}$$

(outcomes a_i and b are conditionally independent given atomic events x_j). Therefore the update rule for the statistical operator ρ is

$$\rho(x_j) \to \rho'(x_j) = \frac{\rho(x_j) p(a_i|x_j)}{\sum_j \rho(x_j) p(a_i|x_j)}. \tag{7.22}$$

The same strategy as Bub uses for projective measurements suggests that the quantum POVM operator E_i is the quantum counterpart of the conditional probability $p(a_i|x_j)$.

This association also appears in the treatment of Jacobs and Fuchs. They both make the connection between quantum measurement and Bayesian updating for POVMs, not just projective measurements, but they focus on the case of efficient POVMs. As we have noted earlier, all projective measurements fall into this category.

Jacobs proceeds by first writing classical measurements in a similar language to generalized quantum measurements. The initial probability vector $p(h)$ is written as a diagonal matrix ρ, with $(\rho)_{hh} = p(h)$, and the distribution $p(d|h)$ is written in a diagonal matrix E_d such that $(E_d)_{hh} = p(d|h)$. Then the conditionalized probability distribution is given by another diagonal matrix ρ_d such that $(\rho_d)_{hh} = p(h|d)$. In this terminology, the classical Bayes rule becomes

$$\rho_d = \frac{E_d \rho}{p_d}. \tag{7.23}$$

To get to quantum measurements, Jacobs suggests relaxing the restriction that the matrices involved must be diagonal, while keeping the restriction that they must be positive [7]. Since in the classical case all the matrices commute, Jacobs points out that a symmetrized version of (7.23) in the non-commuting case can be written as

$$\rho'_d = \frac{\sqrt{E_d}\rho\sqrt{E_d}}{p_d}. \tag{7.24}$$

This establishes that the quantum state-update rule is *a* non-commutative generalization that reduces to classical Bayesian updating when the operators commute. Bub establishes a similar point in claiming that (7.18) is a non-commutative generalization of (7.19). However, to make the stronger claim that the quantum state update *is* Bayesian conditionalization in the non-commutative case, one would have to also show that there is some reason why this, and not other symmetrized versions of (7.23), is the appropriate generalization. In particular, another symmetrized version of (7.23) is

$$\widetilde{\rho}_d = \frac{\rho^{1/2} E_d \rho^{1/2}}{p_d}. \tag{7.25}$$

It is interesting that the quantum state-update rule takes us to a final state ρ'_d that is just a unitary transformation away from $\widetilde{\rho}_d$,

$$\rho'_d = U_d \widetilde{\rho}_d U_d^{\dagger}. \tag{7.26}$$

Fuchs points out that a transition to $\widetilde{\rho}_d$, rather than ρ'_d would correspond exactly to Bayesian updating of the probabilities associated with the states. The initial probabilities associated with ρ for measurements of the POVM $\{E_h\}$ are

$$p(h) = \mathrm{tr}(\rho E_h). \tag{7.27}$$

If the updated state after measuring d were $\widetilde{\rho}_d$, then the new probabilities would be

$$p_d(h) = \mathrm{tr}(\widetilde{\rho}_d E_h). \tag{7.28}$$

It follows from the completeness relation $\sum_d E_d = 1$ that

$$\rho = \sum_d p_d \widetilde{\rho}_d. \tag{7.29}$$

Therefore, $p(h) = \mathrm{tr}(\rho E_h) = \sum_d p_d \mathrm{tr}(\widetilde{\rho}_d E_h)$. Comparing this equation to $p(h) = \sum_d p(d)p(h|d)$ gives

$$p(h|d) = \mathrm{tr}(\widetilde{\rho}_d E_h). \tag{7.30}$$

Therefore, Bayesian updating $p(h) \to p(h|d)$ corresponds to the transition in the state $\rho \to \widetilde{\rho}_d$. However, the quantum state update is not $\rho \to \widetilde{\rho}_d$, but $\rho \to \rho'_d$, where $\rho'_d = U_d \widetilde{\rho}_d U_d^\dagger$, so it involves an "extra unitary" transformation.

Fuchs therefore interprets the quantum measurement as conceptually comprising two parts:

(1) a "refinement" as in the classical case,

$$\rho \overset{d}{\to} \widetilde{\rho}_d; \tag{7.31}$$

(2) a further "mental readjustment" of the observer's beliefs,

$$\widetilde{\rho}_d \overset{U_d}{\to} \rho_d. \tag{7.32}$$

Fuchs says "Quantum measurement is nothing more, and nothing less, than a refinement and a readjustment of one's initial state of belief" ([3], p. 34).

This conceptual breakdown of the measurement applies to efficient measurements only, where there is only one index. An inefficient measurement cannot be regarded as comprising these two steps because the final state

$$\rho'_d = \sum_{d'} \frac{E_{dd'}^{1/2} \rho E_{dd'}^{1/2}}{p_{dd'}} \tag{7.33}$$

is neither a unitary transformation away from a state $\widetilde{\rho}_{dd'} = (1/p_{dd'})\, \rho^{1/2} E_{dd'} \rho^{1/2}$ produced by a refinement of the initial state

$$\rho = \sum_{dd'} p_{dd'} \widetilde{\rho}_{dd'} \tag{7.34}$$

nor is it a unitary transformation away from the state

$$\widetilde{\rho}_d = \sum_{d'} \frac{\rho^{1/2} E_{dd'} \rho^{1/2}}{p_{dd'}}. \tag{7.35}$$

The reason is that the final state is a mixture of states, where each state $\rho_{dd'}$ is a unitary transformation away from $\widetilde{\rho}_{dd'}$. However, we have lost the information about which d' the state is associated with. Therefore, there is no way to apply a single unitary to retrieve the original state. We will refer to the measurements for which the refinement-plus-unitary interpretation is appropriate "generalized Bayesian measurements."

7.5 Quantum measurement and information gain

In this section we will show that the extra unitary transformation appearing in quantum measurement gives rise to an interesting disanalogy between classical and quantum cases in terms of information gain on measurement.

7.5.1 Information gain

There are several different measures of information gain in measurement (e.g., [6, 9, 10]). Here we will be interested in a particular type of information gain that can be expressed either in terms of a majorization relation, or in terms of an entropy inequality. The property of interest concerns how the final uncertainty (averaged over all the possible final states) compares with the uncertainty in the initial state. In particular, one hopes that the uncertainty after a measurement is less than the uncertainty before. We will now give a brief explanation of how this property can be represented using majorization and entropy inequalities.

The intuitive interpretation of the majorization relation $\prec$ is that, if $x \prec y$, then x is more "mixed," "disordered," or "uncertain" than y. It is defined as follows.

Definition 1. *Suppose* $x = (x_1, x_2, \ldots, x_d)$ *and* $y = (y_1, y_2, \ldots, y_d)$ *are two vectors. If we reorder the components of* x *to be in non-increasing order, we denote the new vector by* $x^{\downarrow} = (x_1^{\downarrow}, x_2^{\downarrow}, \ldots, x_d^{\downarrow})$*, with* $x_1^{\downarrow} \geq x_2^{\downarrow} \geq \ldots \geq x_d^{\downarrow}$*. Then we say that* x *is majorized by* y*,* $x \prec y$*, if*

$$\sum_{j=1}^{k} x_j^{\downarrow} \leq \sum_{j=1}^{k} y_j^{\downarrow} \tag{7.36}$$

for $k = 1, \ldots, d-1$*, and equality for* $k = d$*.*

The connection to mixedness or disorder is expressed in the following theorem.[4]

Theorem 1. $x \prec y$ *iff* $x = \sum_j p_j P_j y$ *for some probability distribution* p_j *and permutation matrices* P_j*.*

This means that x is more disordered than y because it can be obtained by applying various permutations to y and then mixing the results.

We can apply the concept of majorization to quantum states via the following definition.

Definition 2. *Suppose* ρ *and* σ *are Hermitian operators. Then we say that* $\rho \prec \sigma$ *if the vector* $\lambda(\rho)$ *of eigenvalues of* ρ *is majorized by the vector* $\lambda(\sigma)$ *of eigenvalues of* σ*,* $\lambda(\rho) \prec \lambda(\sigma)$*.*

[4] For a proof of this theorem, see [11], p. 574.

The majorization relation

$$\lambda(\rho) \prec \sum_d p_d \lambda(\rho'_d) \tag{7.37}$$

therefore expresses the idea that the initial state ρ is more mixed on average than the states ρ'_d comprising the post-measurement ensemble.

The second way to express this idea is in terms of an inequality of entropies. Suppose we have a random variable X which can take values $x_1, \ldots, x_n$ with probabilities $p_1, \ldots, p_n$. The Shannon entropy associated with this probability distribution is

$$H(X) = -\sum_d p_d \log p_d. \tag{7.38}$$

Intuitively the Shannon entropy may be thought of as a measure of our uncertainty about the random variable before we learn what value it takes. Each choice of measurement $\{E_i\}$ on a quantum state ρ produces a random variable corresponding to the outcomes X with probabilities given by $\{p_i = \mathrm{tr}(\rho E_i\}$, and Shannon entropy $H_\rho(X) = -\sum_i p_i \log p_i$. The entropy of the state ρ itself may be defined as

$$S(\rho) = -\mathrm{tr}(\rho \log \rho). \tag{7.39}$$

This quantity is called the "Von Neumann entropy." It is equal to the Shannon entropy of the probability distribution arising from a projective measurement onto the state's eigenvectors. In fact, this measurement produces the lowest Shannon entropy of any projective measurement on the state, and so $S(\rho) \leq H(X)$, where X is the random variable of measurement outcomes for other projective measurements. The von Neumann entropy provides a measure of the uncertainty in a quantum state, just as the Shannon entropy provides a measure of the uncertainty in a probability distribution.

It can be shown that there is a close connection between majorization and entropies. If we have two quantum states ρ and σ such that $\rho \prec \sigma$, then $S(\rho) \geq S(\sigma)$, where $S(\rho)$ is the von Neumann entropy of the state ρ. Therefore the property that the initial state is more mixed than the final state on average in a measurement, given by the majorization relation (7.37), can also be expressed by the entropy inequality[5]

$$S(\rho) \geq \sum_d p_d S(\rho'_d). \tag{7.40}$$

[5] The difference $S(\rho) - \sum_d p_d S(\rho'_d)$ is a quantity of interest in quantum feedback control, where it serves as an indication of the purifying power of the measurement, or the amount of state reduction [9].

(Classically we can define a similar relation, using Shannon entropies in place of von Neumann entropies.) We now have a precise formulation of the information-theoretic property of interest, expressed by either the majorization relation (7.37) or the entropy inequality (7.40). I will refer to this informally as the property of "decreasing our uncertainty" or "increasing our information" about the state on measurement.

7.5.2 Information gain in quantum measurements

Nielsen has proved results that allow us to find necessary and sufficient conditions on the form of a quantum measurement that satisfies the majorization relation (7.37). Nielsen shows that for efficient measurements, relation (7.37) always holds between the initial and final states, i.e., the following theorem holds.

Theorem 2. *Suppose* $\{A_d\}$ *is a set of measurement matrices satisfying* $\sum_d A_d^\dagger A_d = 1$. *Then*

$$\lambda(\rho) \prec \sum_d p_d \lambda(\rho'_d), \tag{7.41}$$

where $\rho'_d = A_d \rho A_d^\dagger / \mathrm{tr}(A_d \rho A_d^\dagger)$ *and* $p_d = \mathrm{tr}(A_d \rho A_d^\dagger)$.

This theorem was proved also by Fuchs and Jacobs [6], using a different method.

In addition, Nielsen showed that there is a partial converse, given by the following theorem.

Theorem 3. *Suppose* ρ *is a density matrix with vector of eigenvalues* λ, *and* ρ'_d *are density matrices with vectors of eigenvalues* λ_d. *Suppose* p_d *are probabilities such that* $\lambda(\rho) \prec \sum_d p_d \lambda_d$. *Then there exist* $\{A_{dd'}\}$ *and a probability distribution* $p_{dd'}$ *such that*

$$\sum_{dd'} A_{dd'}^\dagger A_{dd'} = I, \tag{7.42}$$

$$p_{dd'} \rho'_d = A_{dd'} \rho A_{dd'}^\dagger, \tag{7.43}$$

$$p_d = \sum_{d'} p_{dd'}. \tag{7.44}$$

This is not a full converse, because there are some information-increasing measurements that cannot be expressed with just one index. Nielsen gives the following example. Let $\rho = I/2$ and $\rho' = |0\rangle\langle 0|$. Then $\lambda(\rho) \prec \lambda(\rho')$. Yet there is no one-index measurement that takes ρ to ρ'. However, there is a two-index measurement that does, namely $\{|0\rangle\langle 1|, |0\rangle\langle 0|\}$.

The conditions on the form of the measurement in Theorem 3 are not just a necessary condition for majorization relation (7.37), but are also sufficient. Showing this is a simple extension of the proof of Theorem 4 by Fuchs and Jacobs [6].

Suppose there exist $\{A_{dd'}\}$ and a probability distribution $p_{dd'}$ such that

$$\sum_{dd'} A^\dagger_{dd'} A_{dd'} = I, \tag{7.45}$$

$$p_{dd'}\rho'_d = A_{dd'}\rho A^\dagger_{dd'}, \tag{7.46}$$

$$p_d = \sum_{d'} p_{dd'}. \tag{7.47}$$

Multiplying (7.45) on the left and right by $\rho^{1/2}$ gives

$$\rho = \sum_{dd'} p_{dd'}\widetilde{\rho}_{dd'}, \tag{7.48}$$

where

$$\widetilde{\rho}_{dd'} = \frac{1}{p_{dd'}}\rho^{1/2}A^\dagger_{dd'}A_{dd'}\rho^{1/2}. \tag{7.49}$$

For any convex combination of quantum states ρ_i with probabilities p_i, the majorization relation $\lambda(\rho) \prec \sum_i p_i\lambda(\rho_i)$ holds (see Theorem 1 of [5]). Therefore

$$\lambda(\rho) \prec \sum_{dd'} p_{dd'}\lambda_{dd'}, \tag{7.50}$$

where $\lambda_{dd'} = \lambda(\widetilde{\rho}_{dd'})$. Now, since $X^\dagger X$ and $XX^\dagger$ have the same eigenvalues for any operator X, it follows that $\lambda(\widetilde{\rho}_{dd'}) = \lambda(\rho'_d)$, where

$$\rho'_d = \frac{1}{p_{dd'}}A_{dd'}\rho A^\dagger_{dd'} \tag{7.51}$$

and $\sum_{d'} p_{dd'} = p_d$, and therefore

$$\lambda(\rho) \prec \sum_d p_d\lambda(\rho'_d). \tag{7.52}$$

7.5.3 Generalized Bayesian measurements and information gain

We now show that the necessary and sufficient conditions for measurements satisfying the majorization relation are also necessary and sufficient conditions for the measurement to be a "generalized Bayesian measurement" comprising a refinement and unitary transformation. Thus generalized Bayesian measurements always decrease our uncertainty about the state, and conversely, if we decrease uncertainty, on going from a particular initial state to a set of final states, then there must be

a generalized Bayesian measurement between these states. This is shown in the following theorem.

Suppose there exists a measurement on an initial state ρ and the final states are ρ'_d with probability distribution p_d. If the measurement can be interpreted as a generalized Bayesian updating, then the relation for the initial state

$$\rho = \sum_d p_d U_d \rho'_d U_d^\dagger \tag{7.53}$$

must hold for some unitaries U_d. One refines the initial state to $U_d \rho'_d U_d^\dagger$, and then performs a unitary transformation U_d to obtain the final state ρ'_d. Let B be the condition that there exists a measurement satisfying (7.53).

Let M be the majorization condition

$$\lambda(\rho) \prec \sum_d p_d \lambda(\rho'_d). \tag{7.54}$$

Then we have the following theorem.

Theorem 4.

$$B \Leftrightarrow M. \tag{7.55}$$

Proof. ($\Rightarrow$) Suppose $\rho = \sum_d p_d U_d \rho'_d U_d^\dagger$ for some unitaries U_d. Let $\widetilde{\rho}_d = U_d \rho'_d U_d^\dagger$. Since $\rho = \sum_d p_d \widetilde{\rho}_d$

$$\lambda(\rho) \prec \sum_d p_d \lambda(\widetilde{\rho}_d) = \sum_d p_d \lambda(\rho'_d) \tag{7.56}$$

by Theorem 1 of [5], and $\lambda(\rho'_d) = \lambda(\widetilde{\rho}_d)$.

($\Leftarrow$) Suppose $\lambda(\rho) \prec \sum_d p_d \lambda(\rho'_d)$. Then, by Nielsen's Theorem 2 above, there exist matrices $\{A_{dd'}\}$ and a probability distribution $p_{dd'}$ such that

$$\sum_{dd'} A_{dd'}^\dagger A_{dd'} = I, \tag{7.57}$$

$$p_{dd'} \rho'_d = A_{dd'} \rho A_{dd'}^\dagger, \tag{7.58}$$

$$p_d = \sum_{d'} p_{dd'}. \tag{7.59}$$

The matrix ρ'_d has the same eigenvalues as $\widetilde{\rho}_d = (1/p_{dd'}) \rho^{1/2} A_{dd'}^\dagger A_{dd'} \rho^{1/2}$, therefore there is a unitary U_d, such that

$$\widetilde{\rho}_d = U_d \rho'_d U_d^\dagger, \tag{7.60}$$

and we know from the completeness relation $\sum_{dd'} A^\dagger_{dd'} A_{dd'} = 1$ that

$$\begin{aligned}\rho &= \sum_{dd'} p_{dd'} \widetilde{\rho}_d, \\ &= \sum_d p_d U_d \rho'_d U_d^\dagger. \end{aligned} \tag{7.61}$$

7.5.4 Inefficient measurements and information loss

It follows from this theorem that measurements which do not correspond to generalized Bayesian updating can actually increase our uncertainty about the state.[6] An example is the POVM $E_{11} = \frac{1}{2}|0\rangle\langle 0|$, $E_{12} = \frac{1}{2}|+\rangle\langle +|$, $E_{21} = \frac{1}{2}|1\rangle\langle 1|$, $E_{22} = \frac{1}{2}|-\rangle\langle -|$. Then

$$\begin{aligned}\rho'_1 &= \sqrt{E_{11}}\rho\sqrt{E_{11}} + \sqrt{E_{12}}\rho\sqrt{E_{12}} \\ &= \frac{1}{2}(\langle 0|\rho|0\rangle|0\rangle\langle 0| + \langle +|\rho|+\rangle|+\rangle\langle +|) \end{aligned} \tag{7.62}$$

and

$$\begin{aligned}\rho'_2 &= \sqrt{E_{21}}\rho\sqrt{E_{21}} + \sqrt{E_{22}}\rho\sqrt{E_{22}} \\ &= \frac{1}{2}(\langle 1|\rho|1\rangle|1\rangle\langle 1| + \langle -|\rho|-\rangle|-\rangle\langle -|). \end{aligned} \tag{7.63}$$

Such a measurement can increase our uncertainty.

The possibility of an inefficient measurement that can increase our uncertainty contrasts with the classical situation. In classical measurements, the final probability distribution, averaged over all outcomes, is the same as the initial probability distribution $p(h) = \sum_d p(d)p(h|d)$. This is true even if the measurement is inefficient and throws away some information. Consider making a measurement that gives $\sum_{d'} p(h|d, d')$ for a given outcome d. Then

$$\sum_d p(d) \sum_{d'} p(h|d, d') = p(h). \tag{7.64}$$

Therefore in the classical case, the majorization relation (7.37) expressing information gain is trivially satisfied (associating $\sum_{d'} p(h|d, d')$ with ρ_d and $p(h)$ with ρ). The measurement does not reduce certainty about the final state.

It is possible that (7.37) is not satisfied in the classical case, but only if a more general operation is allowed, where a permutation $\Lambda_{dd'}$ is applied before the coarse-graining. Then the new probability distribution after measuring d is given by

[6] Jacobs also notes that some inefficient measurements can increase uncertainty about the state [12].

$$\frac{1}{p(d)} \sum_{d'} \Lambda_{dd'} p(d, d'|h) p(h) \tag{7.65}$$

and the classical version of inequality (7.40) fails.

Classically, inefficient measurements cannot increase uncertainty about the state, unless extra permutations of this kind are introduced before the coarse-graining. The quantum analog of adding permutations is adding unitary transformations[7]

$$\rho'_d = \sum_{d'} \frac{U_{dd'} \sqrt{E_{dd'}} \rho \sqrt{E_{dd'}} U^\dagger_{dd'}}{p_d}. \tag{7.66}$$

However, the states involved in an inefficient quantum measurement can violate (7.37), even without the introduction of further unitaries. In the example we have just given the final states have the form

$$\rho'_d = \sum_{d'} \frac{\sqrt{E_{dd'}} \rho \sqrt{E_{dd'}}}{p_d}. \tag{7.67}$$

In a quantum measurement, there are already built-in "extra unitaries" that function something like the permutations in the classical case, allowing that, in an inefficient measurement, the uncertainty about the state may increase.

7.6 Conclusion

We have seen that Bub and Fuchs share a subjective interpretation of the quantum probabilities as degrees of belief for an observer. However, Bub argues that the Kolmogorovian probability model needs to be generalized in the quantum case so that probabilities may be defined over a non-Boolean algebra. Fuchs, on the other hand, does not propose that the probability model be generalized. In both approaches, there is a close analogy between Bayesian updating and quantum measurement. However, as Fuchs highlights, there is not a full analogy between Bayesian updating and quantum measurement since an efficient quantum measurement is not just a refinement of the probability distribution, but also involves an extra unitary transformation. We have seen that the measurements which can be interpreted as a Bayesian refinement with extra unitary are exactly those measurements which increase the information about the state. On the other hand, there are some quantum measurements that do not fall in this category and that may lose information. This is because averaging over the built-in unitaries has an effect that is similar to what one gets by applying randomizing permutations classically. Nonetheless, the information-losing measurements have no direct classical analog.

[7] See [4] for a more detailed discussion of the role of permutations and unitary transformations.

References

[1] J. Bub, Von Neumann's projection postulate as a probability conditionalization rule in quantum mechanics, *J. Philos. Logic* **6**, 381–390 (1977).
[2] J. Bub, The measurement problem in quantum mechanics, In *Problems in the Foundations of Physics*, G. Toraldo Di Francia (ed.) (Amsterdam: North-Holland, 1979), pp. 71–121.
[3] C. A. Fuchs, Quantum mechanics as quantum information (and only a little more), quant-ph/0205039 (2002).
[4] K. Jacobs, How do two observers pool their knowledge about a quantum system?, *Quantum Info. Processing* **1** 73 (2002).
[5] M. A. Nielsen, Characterizing mixing and measurement in quantum mechanics, *Phys. Rev. A* 63; 022114 (2001).
[6] C. A. Fuchs and K. Jacobs, Information trade-off relations for finite-strength quantum measurements, *Phys. Rev. A* **63**, 062305 (2001).
[7] R. I. G. Hughes, *The Structure and Interpretation of Quantum Mechanics* (Cambridge, MA: Harvard University Press, 1989).
[8] A. Wilce, Quantum logic and probability theory, in *Stanford Encyclopedia of Philosophy* (2006).
[9] A. C. Doherty, K. Jacobs, and G. Jungman, Information, disturbance, and Hamiltonian quantum feedback control, *Phys. Rev. A* **63**, 062306 (2001).
[10] S. Massar and S. Popescu, How much information can be gained by a quantum measurement?, *Phys. Rev. Lett.* 74; 1259 (1995).
[11] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press, 2000).
[12] K. Jacobs, A bound on the mutual information, and properties of entropy reduction, for quantum channels with inefficient measurements, *J. Math. Phys.* 47, 012102 (2006).

Part III

Quantum information

8

Schumacher information and the philosophy of physics

Armond Duwell

8.1 Introduction

Many philosophers and physicists have expressed great hope that quantum information theory will help us understand the nature of the quantum world. The general problem is that there is no widespread agreement on what quantum information is. Hence, such pronouncements regarding quantum information theory as the savior of the philosophy of physics are hard to evaluate. Much work has been done producing and evaluating concepts of information.[1]

In [7] I have defended and articulated the Schumacher [17] concept of quantum information. Roughly speaking, quantum information is construed as the statistical behavior associated with the measurement of a quantum system. Hence it is a coarse-grained operational description of quantum systems, with no recourse to the fundamental ontological features of quantum systems responsible for such behavior. From this perspective, construing quantum mechanics as a theory of quantum information departs from the traditional interpretive endeavors of philosophers and physicists. The question is whether there is any motivation for taking such a view.

The theorem of Clifton, Bub, and Halvorson (CBH) [5] provides just such a motivation. The theorem guarantees that, if a theory T satisfies certain conditions, there will exist an empirically equivalent C^*-algebraic theory that has a concrete representation in Hilbert space, which it is notoriously difficult to interpret as a constructive or mechanical theory. In such a case, any underlying ontologies philosophers develop that are compatible with T will be undermined by the C^*-algebraic equivalent. Bub [2, 3] suggests in light of this in-principle uncertainty regarding ontology that we re-conceive of quantum mechanics as a theory about quantum information.

[1] See Jozsa [16], Deutsch and Hayden [6], Vaidman [24], Duwell [8–10], Fuchs [11], Brukner and Zeilinger [1], and Timpson [20–22].

Philosophy of Quantum Information and Entanglement, ed. A. Bokulich and G. Jaeger. Published by Cambridge University Press.

Bub fails to provide a concept of quantum information, and his proposal is correspondingly difficult to evaluate. In this paper, I substitute the concept of quantum information I advocate into Bub's proposal and tentatively explore the advantages of such a re-conception of quantum theory. In Section 8.2 the theorem of CBH [5] is sketched, and the epistemological problem the theorem generates is discussed. In Section 8.3 the concept of information advocated in [8] is outlined. In Section 8.4 I advocate a partial re-conception of quantum theory as a theory of quantum information. I suggest that it alleviates the epistemological problem generated by the CBH theorem and suggests interesting new work for philosophers and physicists.

8.2 The CBH theorem

CBH [5], together with the supplementary result of Halvorson [14], characterized quantum theories in terms of information-theoretic constraints.[2] By "quantum theory," CBH mean a C^*-algebra with the following features.

(i) Observables are represented by the self-adjoint operators in a non-commutative algebra, but where observables for separate systems commute.
(ii) States are represented by positive normalized linear functionals where there exist entangled states of space-like separated systems.
(iii) Dynamical changes are represented by completely positive linear maps ([3], p. 7).

CBH proved that C^*-algebras will have these features if and only if they satisfy the following information-theoretic constraints:

(i′) no superluminal information transfer via measurement;
(ii′) no broadcasting; and
(iii′) no bit commitment ([3], pp. 7–8).

Constraint (i′) ensures that no measurement interactions on a system can change the statistics associated with measurements on space-like separated systems. Changing the statistics associated with measurements on a system is requisite for information transfer to be possible. Constraint (ii′) prohibits the cloning of pure and mixed states of quantum systems. Constraint (iii′) prohibits certain types of security protocols. Roughly speaking, the constraint forbids making a secure commitment (1 or 0 for instance) such that the commitment cannot be altered by the one making the commitment, nor so that the commitment is impossible to discover unless revealed by the one that made the commitment.

[2] The complete result will be referred to as the CBH theorem, and is taken to include Halvorson's supplementary work.

Perhaps one of the most important consequences of the theorem is that, given any quantum theory that satisfies the information-theoretic constraints, there will exist an empirically equivalent C^*-algebraic theory. Now, that C^*-algebraic theory will have features (i)–(iii) above and a concrete representation in a Hilbert space where value assignments to observables are severely restricted along the lines of Kochen and Specker (1957), Bell (1964), and Bub and Clifton (1996) [4]. So, for any theory that satisfies the information-theoretic constraints, there exists an empirically equivalent theory that is subject to all the interpretive problems that challenge standard quantum mechanics.

This situation raises delicate epistemological issues. The philosopher of quantum mechanics is challenged typically with taking an abstract formal framework used for representing quantum systems and deciding more or less which features of that formal framework correspond to genuine features of quantum systems and which are mere artifacts, and then has to discern which ontologies are compatible with the framework, i.e., to provide an interpretation of that framework. The ontologies discovered through such a process are typically taken to be the best contenders for the real physical ontology, but what licenses such a conclusion is the amazing empirical success of a theory. The existence of empirical equivalents of a theory challenges such conclusions. As a practical matter, all but trivial empirical equivalents are usually hard to come by. Many competing quantum theories that are on the table now, namely GRWP theories, Bohm's theory, no-collapse theories, etc., are not empirically equivalent, but crucial tests to distinguish between these competing theories are not currently possible. In practice, the problem of in-principle empirical equivalents can often be dismissed, but the CBH theorem essentially forces us to confront the problem. It guarantees the existence of an empirically equivalent theory for any theory that satisfies the information-theoretic constraints. So, for any two theories T and T′, which are both acceptable at a particular time, but are not empirically equivalent to each other, if they satisfy the information-theoretic constraints, there will exist a C^* theory empirically equivalent to theory T, as well as one equivalent to T′.

The issue of empirical equivalence hits philosophers of quantum mechanics hardest with respect to the measurement problem. That is where philosophers attempt to spell out how the ontology compatible with their favored quantum theory is able to recover the phenomena. But, if their favored theory satisfies the information-theoretic constraints, there will be an empirically equivalent rival that can make the right predictions, but without postulating any extra structure or ontology, as for example Bohm's theory does with definite values for position as well as a new field. Since the theories are empirically equivalent, there is no empirical evidence that will decide one is better than another. So, the CBH result seems to give philosophers of quantum mechanics a fundamental tradeoff: postulate the

extra structure that secures our experience but without any empirically detectable consequences, or forget the extra structure at the expense of having the appearances remain mysterious.

Bub [3, 4] suggests that the rational epistemological stance is to regard all constructive theories (those that postulate extra structure) as unacceptable. According to Bub, measurement devices are to be treated as black boxes that simply produce a sequence of outcomes according to a probability distribution, which is tantamount to treating measurement devices as Shannon information sources. Hence, Bub advocates re-conceiving quantum mechanics as a theory about the representation and manipulation of physical systems as sources of information, as opposed to a theory about the fundamental constituents of the world, i.e., the fundamental ontology. Physicists and philosophers should forget about articulating an underlying ontology compatible with the observable phenomena in favor of a predictively adequate representation of them. Hence, he is advocating a new role for physics, which many think is not only a great departure from what physics has done, but also from what physics ought to do.

Against Bub, one might argue that physics is in the business not only of producing the right predictions about the world, but also of explaining something. While, on predictive grounds alone, the CBH theorem might make us question the credentials of one theory over another, given C^*-algebraic equivalents' seeming lack of explanatory resources concerning the phenomena, one might easily dismiss them, just as Bub dismisses the theories that postulate "extra" structure.

What constitutes an explanation is in some sense in the eye of the beholder, and that will be discussed below, but there is a fundamental flaw with Bub's proposal. As a general metaphysical tenet, there are physical features of systems that give rise to the dispositions a system has to behave in certain ways in measurement contexts. It would seem a strange creative limitation to not allow physicists and philosophers to explore the underlying ontology. The only way this can be somewhat justified is if the assumptions of the CBH theorem hold, and thus the existence of a C^*-algebraic theory empirically equivalent to a successful constructive theory is guaranteed. That said, it is a completely open question whether the constraints do actually hold. So, the view Bub advocates seems not only too restrictive, but also epistemologically unjustifiable given the uncertain nature of the constraints.[3]

Between these two extreme positions lies interesting middle ground that respects conflicting intuitions regarding the epistemological problem created by the CBH theorem. It involves re-conceiving quantum mechanics as a theory of information, but is an approach that regards foundational work on developing candidates for the

[3] See Duwell [8], Timpson [22], Spekkens [19], Smolin [18], Halvorson and Bub [15], and Halvorson [13] for discussions of the assumptions of the CBH theorem.

fundamental ontology that secures the phenomena as essential to that work. More will be said about this new conception after an appropriate concept of quantum information has been articulated.

8.3 Quantum information

In 1995 Schumacher developed a concept of quantum information that was quite analogous to Shannon's concept of information [17]. The concept of quantum information was developed in the context of a quantum communication system. A quantum communication system consists of a quantum source, which produces a sequence of quantum systems, which are individually in pure states or part of a larger system in a pure state according to a probability distribution; a quantum compression device, which operates on the system produced by the source to minimize the resources required to communicate quantum information; the quantum signal, which is the connection between the two wings of the communication system; and a decompression device, which reproduces the quantum message from the quantum signal.

Timpson [22] has developed a very useful characterization of communication systems utilizing the well-known type/token distinction. Timpson characterizes information sources as producing tokens of different types. The goal of communication is to reproduce tokens of the types that the information source produced tokens of, at the opposite end of the communication system.

Similarly to Shannon's theory, where the Shannon coding theorem provides part of the interpretation of Shannon information, the Schumacher coding theorem provides part of the interpretation of Schumacher information. Schumacher was able to show that the von Neumann entropy associated with a quantum source with average density operator ρ, $S(\rho)$, quantifies the resources necessary and sufficient to reproduce quantum information. In particular, it quantifies the minimal dimension of a Hilbert space requisite to represent the quantum information associated with the system produced by the source.

The minimal resources requisite for communication depend on the criterion of success that is adopted for the communication system. The success criterion that is advocated here is that the entanglement fidelity for the transmission process approaches unity as the length of the sequence produced by the source goes to infinity. The entanglement fidelity of a process is a measure of how well preserved quantum information associated with a system is, as the system undergoes the process. The entanglement fidelity has an operational interpretation as the probability that the initial system and the system after the process will pass a test of sameness. "Sameness" in this context is interpreted probabilistically. Two systems are

considered the same if they behave identically from a statistical point of view in all measurement contexts. What is crucial is that these measurement contexts include not only measurements on the individual systems that compose the sequence, but measurements on parts of the sequence, and even on the sequence with other systems. That means that any entanglement that the initial sequence has, either between the systems that compose the sequence or as part of a larger system, will be preserved if the entanglement fidelity of the process is high.

The descriptive resources are in place to describe what the Schumacher concept of quantum information is. It is that which is quantified by the von Neumann entropy, what I call the *quantum quantity-information*, in order to distinguish it from that which is required to be reproduced for successful communication, what I call the *quantum type-information*. Hence the Schumacher concept breaks into two subconcepts. The quantum-quantity information is simply the minimal resources required for successful communication. The quantum type-information is the statistical behavior of the system produced by the source in all measurement contexts.

It deserves emphasis that the quantum quantity-information is not a quantity that somehow inheres in the quantum signal. It is a quantity that describes the quantum information source, not the particular sequences produced by the source. For a quick verification, note that the von Neumann entropy associated with a source will generally be non-vanishing even when the source produces only systems that are individually in pure states. Hence the quantum quantity-information is not a quantity of substance associated with some particular quantum system whose progress may be tracked through a communication system.

The quantum type-information is the *type* of token produced by the quantum information source, not the token, and hence one cannot talk of the location of the quantum type-information because it is an abstract entity. One can, of course, talk of the location of tokens of particular quantum type-information or instantiations of quantum type-information. It should be noted that quantum information on this analysis is not some kind of substance, unlike in some other concepts of information [16].

It is quite obvious that the quantum type-information can be considered quite independently of quantum quantity-information, and this independence allows the Schumacher conception of quantum information to make contact with Bub's re-conception of quantum mechanics as a theory of quantum information.

8.4 Re-conceiving quantum mechanics

In this section, a particular re-conception is advocated, which deviates from Bub's extreme proposal but nonetheless respects the epistemological problem facing philosophers of quantum mechanics if the assumptions of the CBH theorem hold. I

will advocate a partial re-conception of quantum mechanics as a theory of quantum information. I will discuss the advantages and debits of such a proposal.

The concept of quantum information elaborated has obvious connections to Bub's view of quantum mechanics in light of the CBH theorem. Bub suggests that we ought to view measurement devices as black boxes that produce results according to a probability distribution. Quantum mechanics is to be re-conceived as a theory about representation and manipulation of quantum systems as sources of information. Reconstrued using the concept of information articulated above, quantum mechanics should be conceived of as a theory about the representation and manipulation of quantum information. In terms of the terminology above, quantum measurement devices are those instruments which reveal the properties of a token of particular quantum type-information.[4]

The question many will pose given this re-conception of quantum theory is "Why would it be of any interest, especially to philosophers of physics, to avoid trying to describe the fundamental ontology compatible with our best physical theory in favor of the more coarse-grained quantum information-theoretic approach?" First off, that question sets up a false dichotomy. Unlike Bub, I am not advocating a reorientation of the subject matter of physics, or of philosophy for that matter. Exploring the information-theoretic approach to quantum theories does not entail giving up traditional foundational work. A better question is "What do we gain from considering the information-theoretic perspective in addition to traditional foundational concerns?"

When a philosopher or physicist provides an explanation of a phenomenon, they typically refer to a constructive theory. A mere derivation of a result from standard quantum mechanics is seen as explanatorally vacuous. For example, standard quantum mechanics predicts the right correlations in EPR experiments, but hardly anyone would admit that it explains anything. Similar goes for any C^*-algebraic theories that CBH consider. Constructive theories of quantum phenomena are turned to in order to provide an explanation. Now, one of the traditional desiderata of explanations is that the explananda are true. Well, what confidence should be placed in one of these constructive explanations of a quantum phenomenon in this epistemological situation? Very little, I think. As mentioned above, there are several empirically inequivalent quantum theories that are currently empirically indistinguishable, and some that are empirically equivalent. This is where the quantum information perspective may be very useful.

Currently, all empirically acceptable quantum theories share a wide range of predictions within the bounds of current experimental abilities. Hence, the behavior of

[4] Of course, we know that generally no measurement can reveal the identity of the quantum type the measured system is a token of.

quantum information is shared between all of these currently acceptable theories at this restricted level. This is especially obvious from the operational definition of quantum type-information. Hence quantum information provides a unifying thread to these theories. From that perspective, they are equivalent. Perhaps this equivalence can be used to provide explanations of quantum phenomena utilizing true desiderata without depending on the constructive details of the phenomena. The laws of quantum information supervene on the constructive facts or laws (depending on your metaphysical predilections) and may provide explanations that avoid the problems generated by our current epistemological situation.

Consider an analogous case of explanations of thermal phenomena utilizing thermodynamics instead of statistical mechanics. Suppose that one wants to explain what happens when one throws a hot rock into a pool of cold water. A constructive explanation in terms of the underlying statistical physics is completely intractable, and hardly illuminating. By appealing to the thermodynamic laws, which presumably supervene on the underlying statistical-mechanical behavior of the system, one can make quick work of the phenomenon. The model of explanation that it seems must be appealed to in this case is the unification model. The thermodynamic laws unify a wide range of phenomena under a few simple principles. Similarly, perhaps the laws of quantum information can unify a wide range of quantum phenomena under a few simple principles.

What are the laws of quantum information? Well, no one knows right now, but it is a fascinating and burgeoning area of research. For a taste, consider some recent developments on extensions of the no-cloning theorem. It is well known that quantum states cannot be cloned, hence quantum type-information cannot be cloned. But that is true only if we want perfect clones. Recent work has established how well quantum type-information can be distributed.

Consider a quantum cloning machine that begins the cloning process with N qubits all in the same arbitrary state, i.e., $|\psi\rangle = \sin(\theta/2)\exp(\mathrm{i}\phi)|0\rangle + \cos(\theta/2)|1\rangle$, where the distribution over the states is flat, and is required to produce $M > N$ clones that are all in the same state. To describe how effective the cloning process is, one must use a concept of fidelity. The average fidelity of clones, which is an average over all of the various initial states, is given by

$$\mathcal{F} = \int \mathrm{d}\Omega \langle\psi|\rho_{\mathrm{out}}|\psi\rangle,$$

where $\mathrm{d}\Omega = \int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\theta \sin(\theta)/(4\pi)$. Gisin and Massar [12] have shown that one can produce clones with fidelity

$$\mathcal{F} = \frac{M(N+1)+N}{M(N+2)}.$$

Making a clone of a single system is effective. Choosing $M = 2$, the fidelity of the clones is 5/6. Cloning from N to $N + 1$ is quite effective. The fidelity of the clones tends to unity as N grows. Making clones from a single qubit is surprisingly effective. The fidelity of the 1-to-M cloner goes to 2/3 as M becomes large. One might have expected that when one tries to distribute quantum information as widely as possible the average state of the qubits would bear no relation to the initial state. This example gives one a flavor of the laws of quantum information distribution.

8.5 Conclusion

As remarked at the beginning of this chapter, philosophers and physicists have expressed great hope that quantum information theory will offer great insight into the nature of quantum physics. Whether that is true or not remains to be decided. What is certainly true, though, is that quantum information theory seems to offer interesting prospects for dealing with the epistemological problems of currently acceptable quantum theories, as well as the epistemological problem created by the CBH theorem. In particular, it has been suggested that the concept of quantum information above, coupled with the laws of distribution of quantum information, may provide explanations of quantum phenomena that do not commit one to any particular quantum theory or interpretation thereof. The scope of explanatory power that quantum information theory has to offer remains to be seen. Minimally I hope that quantum information theory is seen by philosophers as a potentially useful philosophical resource, not simply an instrument of application of quantum mechanics.

References

[1] C. Brukner and A. Zeilinger, Conceptual inadequacy of the Shannon information in quantum measurements, *Phys. Rev. A*, **63**, 1–10 (2001).
[2] J. Bub, Why the quantum, quant-ph/0402149 (2004a).
[3] J. Bub, Quantum mechanics is about quantum information, quant-ph/0408020 (2004b).
[4] J. Bub and R. Clifton, A uniqueness theorem for "no collapse" interpretations of quantum mechanics, *Stud. Hist. Philos. Mod. Phys.* **27**, 181–219 (1996).
[5] R. Clifton, J. Bub, and H. Halvorson, Characterizing quantum theory in terms of information-theoretic constraints, *Foundations Phys.* **33**, 1561–1591 (2003).
[6] D. Deutsch and P. Hayden, Information flow in entangled quantum subsystems, *Proc. Roy. Soc. London A* **456**, 1759–1774 (2000).
[7] A. Duwell, Quantum information does exist, *Stud. Hist. Philos. Mod. Phys.* **39**, 195–216 (2008).
[8] A. Duwell, Re-conceiving quantum theory in terms of information-theoretic constraints, *Stud. Hist. Philos. Mod. Phys.* **38**, 181–201 (2007).

[9] A. Duwell, How to teach an old dog new tricks: quantum information, quantum computing, and the philosophy of physics, Ph.D. thesis, University of Pittsburgh, http://etd.library.pitt.edu/ETD/available/etd-12082004-025939/ (2004).
[10] A. Duwell, Quantum information does not exist, *Stud. Hist. Philos. Mod. Phys.* **34**, 479–499 (2003).
[11] C. Fuchs, Quantum mechanics as quantum information, mostly, *J. Mod. Opt.* **50**, 987–1023 (2002).
[12] N. Gisin and S. Massar, Optimal quantum cloning machines, *Phys. Rev. A* **79**, 2153–2156 (1997).
[13] H. Halvorson, On information-theoretic characterizations of physical theories, *Stud. Hist. Philos. Mod. Phys.* **35**, 277–293 (2004).
[14] H. Halvorson, Generalization of the Hughston–Jozsa–Wooters theorem to hyperfinite von Neumann algebras, quant-ph/0310101 (2003).
[15] H. Halvorson and J. Bub, Can quantum cryptography imply quantum mechanics: a reply to Smolin, quant-ph/0311065 (2003).
[16] R. Jozsa, Quantum information and its properties, in *Introduction to Quantum Computation and Information*, H.-K. Lo, S. Popescu, and T. Spiller (eds.) (Singapore: World Scientific, 1998), pp. 49–75.
[17] B. Schumacher, Quantum coding, *Phys. Rev. A* **51**, 2738–2747 (1995).
[18] J. Smolin, Can quantum cryptography imply quantum mechanics?, quant-ph/0310067 (2003).
[19] R. Spekkens, In defense of the epistemic view of quantum states: a toy theory, quant-ph/0401052 (2004).
[20] C. G. Timpson, The grammar of teleportation, *Brit. J. Philos. Sci.* **57**, 587–621 (2006).
[21] C. G. Timpson, Nonlocality and information flow: the approach of Deutsch and Hayden, *Foundations Phys.* **35**, 313–343 (2005).
[22] C. G. Timpson, Quantum information theory and the foundations of quantum mechanics, Ph.D. dissertation, University of Oxford, quant-ph/0412063 (2004).
[23] C. G. Timpson, On the supposed conceptual inadequacy of the Shannon information in quantum mechanics, *Stud. Hist. Philos. Mod. Phys.* **34**, 441–468 (2003).
[24] L. Vaidman, On the paradoxical aspects of new quantum experiments, in *PSA 1994 Philosophy of Science Association Biennial Meeting*, D. Hull, M. Forves, and R. Burian (eds.) (East Lansing, MI: Philosophy of Science Association, 1994), pp. 211–217.

9

From physics to information theory and back

Wayne C. Myrvold

9.1 Introduction

Quantum information theory is the study of how the peculiar features of quantum mechanics can be exploited for the purposes of information processing and transmission. A central theme of such a study is the ways in which quantum mechanics opens up possibilities that go beyond what can be achieved classically. This has in turn led to a renewed interest in, and a new perspective on, the differences between the classical and the quantum. Although much of the work along these lines has been motivated by quantum information theory – and some of it has been motivated by the conviction that quantum theory is essentially about possibilities of information processing and transmission – the results obtained, and the frameworks developed, have interest even for those of us who are not of that conviction. Indeed, much of the recent work echoes, and builds upon, work that predates the inception of quantum information theory. The significance of such work extends beyond the setting of quantum information theory; the work done on distinguishing the quantum from the classical in the context of frameworks that embrace both is something worthy of the attention of anyone interested in the foundational issues surrounding quantum theory.

One of the striking features of quantum mechanics lies in its probabilistic character. A quantum state yields, not a definite prediction of the outcome of an experiment, but a probability measure on the space of possible outcomes. Of course, probabilities occur also in a classical context. In this context they have to do with situations in which the experimenter does not have complete control over the classical state to be prepared, and, as a consequence, we do not have complete knowledge of the classical state of the system subjected to the preparation procedure. The question which arises, therefore, is whether quantum probabilities can be construed as being like this. One way of framing this question is in terms of hidden variables: ought we to think of quantum-mechanical pure states as being probabilistic

Philosophy of Quantum Information and Entanglement, ed. A. Bokulich and G. Jaeger. Published by Cambridge University Press.

mixtures of states of a more encompassing theory, whose pure states would ascribe definite values to all variables? It is interesting to ask which of the peculiar features of quantum mechanics are traceable to ineliminable statistical dispersion in its states. Such features will be reproducible in an essentially classical theory with suitable restrictions on possibilities of state preparation.

Some of the recent work in quantum information theory has shown that some of the features of quantum mechanics that one might be inclined to think of as peculiarly quantum can, indeed, be recovered from a theory in which the preparable states are probabilistic mixtures of such classical states. The no-cloning theorem is a case in point. Arbitrary pairs of quantum states cannot be cloned; those pairs that can be cloned are orthogonal pairs. Though originally formulated in the context of quantum mechanics, it admits of a formulation applicable to classical mixed states, which are probability measures over a classical phase space. A pair of classical states is orthogonal iff the probability measures have disjoint support, and it can be shown that clonable pairs are orthogonal in this sense. The similarity between these two theorems suggests that they are special cases of a more general theorem, and this is indeed the case. Implicit in the proof of Lemma 3 of Clifton, Bub, and Halvorson [11] is a proof that a pair of pure states of a C^*-algebra are clonable iff they are orthogonal. Barnum, Barrett, Leifer, and Wilce [5] prove a theorem of much greater generality. Within the convex-sets framework (see Section 9.4), they prove that a finite set of states on a compact, finite-dimensional state space is clonable if and only if it is a jointly distinguishable set of states.

Some, notably Fuchs [13] and Spekkens [29, 30], have found in this fact – that some phenomena that might be thought to be distinctively quantum can be reproduced in a classical theory by imposing restrictions on state preparation – encouragement for the view that quantum probabilities are just like classical probabilities, epistemic probabilities bound up with limitations on state preparation. A natural alternative is to conclude that those features that can be reproduced in an essentially classical setting ought not to have been considered distinctively quantum in the first place. This is, I think, the right lesson to draw. If this is right, then we must seek deep distinctions between the classical and the quantum elsewhere. To anticipate a conclusion to be drawn below, a case can be made that Schrödinger [27] was right to locate the essential difference between the classical and the quantum in the treatment of combined systems. However, Schrödinger's conclusion that it is entanglement that distinguishes the quantum from the classical requires qualification – as we shall see.

In this paper, I will compare and contrast two approaches to the construction of neutral frameworks in which theories can be compared. The first is the algebraic framework, which begins with an algebra, among the elements of which are included the observables of the theory. The second is the operational approach,

which motivates the introduction of the convex-sets framework. I will end with a discussion of one particular model, the toy theory devised by Rob Spekkens [29, 30], which, with minor modifications, fits neatly within the convex-sets framework and displays, in an elegant manner, some of the similarities and differences between classical and quantum theories.

9.2 Algebraic frameworks

Clifton, Bub, and Halvorson [11] (henceforth CBH) undertook the task of characterizing quantum mechanics in terms of information-theoretic constraints. They adopted a framework in which a physical theory is associated with a C^*-algebra, the self-adjoint elements of which represent the bounded observables of the theory. For the definition of a C^*-algebra, see the appendix to this chapter; for our purposes it suffices to know that the set of all bounded operators on a Hilbert space is a C^*-algebra, as is any sub-algebra of these that is closed under the operation of taking adjoints, and is complete in the operator norm. Moreover, the set of all bounded, continuous complex-valued functions on a classical phase space is a C^*-algebra, as is the set of all bounded, measurable complex-valued functions on a classical phase space. Thus, classical mechanics also admits of a C^*-algebra representation, as was shown by Koopman in 1931. The difference between the quantum case and the classical case is that, in the classical case, the algebra is Abelian.

A *state* on a C^*-algebra $\mathfrak{A}$ is a positive linear functional $\rho\colon \mathfrak{A} \to \mathbb{C}$, normalized so that $\omega(I) = 1$. For self-adjoint A, the number $\omega(A)$ is to be interpreted as the expectation value of the observable corresponding to A, in state ω. The set of states is a convex set: for any states ρ, σ, and any real $\lambda \in (0, 1)$, the functional defined by

$$\omega(A) = \lambda\, \rho(A) + (1 - \lambda)\, \sigma(A)$$

is also a state, a mixture of ρ and σ. A state that is not a mixture of any two distinct states is called *pure*. General state evolution is represented by completely positive norm-preserving linear maps, also known as non-selective operations.

A state is *dispersion-free* iff $\rho(A^2) = \rho(A)^2$ for all self-adjoint A. Any dispersion-free state is pure. It can be shown that a C^*-algebra is Abelian iff all its pure states are dispersion-free. It can be shown also that a theory involving an Abelian C^*-algebra admits of an essentially classical representation, in which the states are probability distributions on the set of its pure states. Thus, within the C^*-algebraic framework, the classical theories are those whose algebras are Abelian.

Within this framework, CBH characterize quantum theory via three properties of the algebra.

(i) Algebras associated with distinct physical systems commute.
(ii) Any individual system's algebra of observables is non-commutative.
(iii) Space-like separated systems at least sometimes occupy entangled states.

As CBH ([11], p. 1570) point out, the first two conditions entail that the state space of a composite system contains non-locally entangled states. What the third requirement is meant to do is to guarantee that these states are physically accessible. Thus, CBH allow for theories in which the set of preparable states is a proper subset of the full state space of the algebra of observables. If the set of preparable states is taken to be the full set of states of a C^*-algebra, then the third condition is redundant.

Though CBH describe their conditions as "definitive of what it means to be a quantum theory in the most general sense" ([11], p. 1563), it should be noted that these three properties do not suffice to characterize quantum mechanics. In a quantum theory, there are no states that are dispersion-free in all observables. This is not entailed by CBH's conditions, as can be shown by the following simple example. Let A and B be two separated systems, and associate with each of these the algebra $M(\mathbb{C}) \oplus M(\mathbb{C}^2)$ – that is, the algebra of 3×3 complex matrices of the form

$$\begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & \gamma \\ 0 & \delta & \epsilon \end{pmatrix}.$$

Associate with the composite system the algebra that is the tensor product of the algebras associated with A and B.

All three of CBH's conditions are satisfied. However, unlike in either quantum mechanics or classical mechanics, some of the pure states are dispersion-free and some are not. The state corresponding to the vector

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

is an eigenvector of every observable. To ensure that the state space of our theory is quantum mechanical, some additional condition is needed. Plausible candidates are symmetry conditions; one might impose, for example, the condition that, for any pair of pure states ρ, σ on $\mathfrak{A}$ (or $\mathfrak{B}$), there is an automorphism of the algebra $\mathfrak{A}$ taking ρ into σ. This condition would entail that either all pure states are dispersion-free (in which case the algebra is Abelian, and the state space is a classical simplex), or that none are (as in the quantum case).

The virtues of the C^*-algebraic formulation are that there is a rich and well-worked-out theory of C^*-algebras, and that both classical and quantum

theories – including quantum mechanics and quantum field theories – are readily formulated in such terms. This rich theory comes at a price, however. The assumption that the set of observables be the self-adjoint part of a C^*-algebra requires that sums and products of observables be well defined even when the observables are incompatible ones. When self-adjoint operators A, B fail to commute, the product AB will not correspond to any observable; nevertheless, our definition of a state requires that states assign numbers to such products, numbers that are not interpretable as expectation values of the results of measurement. The embedding of our observables into an algebra imposes non-trivial algebraic relations between expectation values assigned to observables.

It would be difficult to argue – or at least, at this point nobody knows how to argue – that any plausible physical theory would admit of a C^*-algebraic formulation. If we further require that the set of preparable states be the full state space of some C^*-algebra, then the C^*-algebraic framework becomes decidedly too restrictive. Halvorson [16] discusses the case of a theory that he calls the Schrödinger theory, in which elementary systems are like quantum systems, but in which entangled states decay into mixtures when the systems are separated. Such a theory is locally quantum, but admits of no Bell-inequality-violating correlations. Such a theory was suggested by Schrödinger [27], who pointed out that at the time there was little in the way of experimental evidence of non-locally entangled states. Though there is now abundant evidence of non-local entanglement, it does not seem that the Schrödinger theory is one that could, or should, have been ruled out in advance of experiment. Moreover, we would like a framework in which we can consider some admittedly artificial constructions, such as Spekkens' toy theory, discussed in Section 9.5. As Halvorson [16] has shown, the state space of the Spekkens theory is not the state space of a C^*-algebra.

Within the algebraic approach, one can also consider weakening of the algebraic assumptions. One can, for example, consider Jordan–Banach (JB) algebras, or Segal algebras (see Halvorson [16] for definition and discussion), both of which contain C^*-algebras as special cases. Seeking a further widening of the algebraic framework, with constraints limited to those that, arguably, any reasonable physical theory must share, one is led naturally to effect algebras.[1]

An effect algebra is meant to represent the set of yes–no tests that can be performed on a physical system. It contains distinguished elements 0 and u (the unit element), representing the two trivial tests whose outcomes, independently of the state, are "no" and "yes," respectively. There is a *partial* operation $\oplus$. $a \oplus b$ is

[1] The notion of an effect algebra has occurred in the writings of many authors working on the foundations of quantum mechanics. The presentation here is based on that of Beltrametti and Bugajski [10]. It should be pointed out that, despite the name, an effect algebra is not an algebra: mutiplication is not defined, and addition is a partial operation.

defined when a and b can represent alternative outcomes of a single experiment. The partial operation $\oplus$ is assumed to be symmetric and associative. Moreover, it is assumed that, for each effect a, there is a complementary effect $\bar{a}$ such that $a \oplus \bar{a} = u$. As a final condition, $a \oplus u$ is defined only when $a = 0$. A *resolution of the identity* is a set $\{a_i, \ldots, a_n\}$ such that

$$\sum_{i=1}^{n} a_i = u.$$

Associated with an n-outcome experiment is a resolution of the identity representing its possible outcomes.

An element a of a C^*-algebra $\mathfrak{A}$ is said to be *positive* iff $a = b^*b$ for some $b \in \mathfrak{A}$. We can define a partial order $\leq$ on $\mathfrak{A}$ by $a \leq b$ iff $a + c = b$ for some positive c. The set of positive elements a of $\mathfrak{A}$ such that $0 \leq a \leq I$ forms an effect algebra, with $a \oplus b$ defined to be $a + b$ when $0 \leq a + b \leq I$, and undefined otherwise.

The conditions defining an effect algebra, unlike those defining C^*-algebras or JB algebras, have clear significance in terms of their intended association with experiments. This gives us a broad framework within which we add conditions to delimit interesting classes of physical theories.

Though restricting ourselves to the C^*-algebraic framework excludes some theories that we might want to consider, such a framework nonetheless has a role to play. As mentioned, the C^* framework is broad enough to include both classical and quantum theory. Moreover, it lends itself readily to hybrid theories, since it permits hybrid systems, composite systems having both classical and quantum subsystems. This is useful for considering what can be done by a combination of classical and quantum information processing. Moreover, a deep theorem by Alfsen, Hanche-Olsen, and Shultz, discussed in Section 9.4, shows that the C^*-assumption entails that the simplest systems are equivalent either to a classical bit or to a qubit. This is an indication that the theory is too restrictive to be regarded as a *general* framework for physical theories, and suggests rather that it is the minimally general framework that is broad enough to include both the classical and the quantum.

9.3 The operational approach

Lucien Hardy ([19, 20]) sets out to characterize quantum mechanics within a framework whose basic concepts are framed in terms of performable operations. Hardy invites the reader to imagine concrete devices (state-preparation devices, transformation devices, and measurement devices), and characterizes quantum mechanics in terms of relations between possibilities of state preparation, transformation, and measurement. States, in this approach, are associated with preparation

devices, and are regarded as compendia of probabilities regarding the outcomes of any measurement that can be performed. Similar discussions can be found in Holevo [21], Ludwig [24], and D'Ariano [12].

To take such concepts as basic may seem to smack of an operationalism that eschews on principle any talk of a reality not directly accessible to observation and manipulation. This is not, however, a necessary concomitant of this approach. If one wishes to construct a framework capable of embracing as wide an array as possible of physical theories, it can be a useful strategy to restrict one's attention to the features that *any* physical theory might be expected to have. Whatever else it might do, a physical theory should specify a set of possible states of a physical system, and say something about operations that can be performed on systems and experiments that can be done. The states, however (or irrespective of whether) they may be conceived ontologically, ought, in the context of the physical theory, to yield probabilities for outcomes of any experiment that might be performed. The operationalism associated with this approach can be thought of as a *methodological operationalism*. Someone who takes as the goal of theorizing an account of a reality existing independently of us and our experimental manipulations might nonetheless wish to adopt a framework that begged no questions about what that reality is like. The framework itself is neutral between an attitude on which operational concepts are the only meaningful ones, and one that seeks to use them as a springboard for theorizing about the nature of physical reality.

It is also neutral between an attitude that rests content with taking the concepts of preparation, transformation, and experiment as primitives, and does not seek to explain them within the theory to be constructed, and an attitude that seeks to "close the circle," to borrow a phrase from Abner Shimony [28], by construing the experimental apparatus as among the physical systems to be dealt with by the theory, and the processes by which states are prepared and experimental results obtained to be among the dynamical evolutions allowed by the theory. It is the latter, of course, that has proved problematic in connection with quantum mechanics!

States are associated with preparation procedures, and yield probabilities over the results of experiments, given the preparation procedure. These probabilities might or might not represent the epistemic state of some human agent. An agent might be in doubt about what probabilities are most appropriate to associate with a given preparation procedure. She might, for example, be uncertain whether a given combination of a preparation and a two-outcome experiment yielded equal probabilities for both outcomes. The conjecture that the probabilities are equal can be subjected to test; by repetition our agent can satisfy herself, to any degree of precision and confidence required, about what probabilities she ought to associate with a given combination of preparation and experiment.

Proponents of a subjectivist interpretation of probability in quantum mechanics sometimes assert that it is nonsensical to speak of an unknown quantum state (see, e.g., Fuchs and Schack [14]). Indeed, if the probabilities associated with a state had to be epistemic probabilities, then a state could be unknown only if an agent were ignorant of her own state of mind. However, it is surely not nonsensical to say that a system has been subjected to some state-preparation procedure, though no-one knows what procedure was applied. In quantum information theory situations are routinely invoked in which one agent has prepared a state, and another has to guess what state it is. This situation would not be fundamentally different if it were automated; we can imagine a machine subjecting a system to some preparation procedure, making a record of which procedure was applied.

The operational approach gives us the following picture: we have a set of states, associated with preparation procedures; a set of possible transformations of the states; and a set of possible experiments that can be performed. The mathematical framework that lends itself naturally to this picture is the convex-sets framework, to be discussed in the next section.

9.4 The convex-set approach

Given any two preparation procedures ρ, σ and any real number $\lambda \in (0, 1)$, there will be a third, according to which one utilizes either procedure ρ or σ, with probabilities λ and $1-\lambda$, respectively. Thus, we expect the set of states of any reasonable physical theory to be convex: it will contain all mixtures of all states that it contains.

The convex-set approach starts with a convex state space Ω.[2] Extremal points of this state space – that is, states that cannot be expressed as non-trivial mixtures of other states – are called *pure*. An *affine linear functional* on Ω is a mapping $a\colon \Omega \to \mathbb{R}$ that respects mixtures; that is,

$$a(\lambda\rho + (1-\lambda)\sigma) = \lambda a(\rho) + (1-\lambda)a(\sigma),$$

for all $\rho, \sigma \in \Omega$, $\lambda \in (0, 1)$. Among affine linear functionals are the constant functionals, taking on the same value on every state; we single out the unit functional u and the zero functional 0. There is a natural partial ordering on $\mathcal{A}(\Omega)$, the set of affine linear functionals: $a \leq b$ iff $a(\omega) \leq b(\omega)$ for all $\omega \in \Omega$. If $0 \leq a \leq u$, a is called an *effect*. The set of effects on a state space Ω will be denoted by $E(\Omega)$. The notion of an effect generalizes the notion, which is familiar from the quantum context, of an effect as a positive operator with spectrum in [0, 1]. A resolution of the

[2] The presentation of the convex-set framework presented here is heavily indebted to the presentations in Beltrametti and Bugajski [10], and in Barnum *et al.* [5].

identity u in terms of effects generalizes the notion of a positive operator-valued measure (POVM).

The set of effects has a natural structure as an effect algebra in the sense of Section 9.2, and, conversely, the set of all probability measures on an effect algebra forms a convex state space. Relations between these frameworks are further discussed in Beltrametti and Bugajski [10].

There is a natural affine structure on the set of effects: for any effects a, b, and any $\lambda \in (0, 1)$, the functional c defined by

$$c(\omega) = \lambda a(\omega) + (1 - \lambda)b(\omega)$$

is also an effect. Thus, we can distinguish between effects that can be written as non-trivial mixtures of other effects and those that cannot. The former are called *mixed* effects, the latter, *pure*. A resolution of the identity in terms of pure effects is what corresponds, in the context, to a quantum projection-valued measure (PVM).

Effects map states to probabilities. Equivalently, we can think of states as mapping effects to probabilities. For any state $\omega \in \Omega$, there is an affine linear mapping $\hat{\omega}\colon E(\Omega) \to \mathbb{R}$ defined by

$$\hat{\omega}(a) = a(\omega).$$

It will be convenient, when defining the tensor-product space, to work with these dual states, which map effects to probabilities.

To specify an observable A, one specifies a measure space $\langle \mathrm{sp}(A), \mathcal{M}_A \rangle$ and an affine mapping α_A from states to probability measures on $\langle \mathrm{sp}(A), \mathcal{M}_A \rangle$. The set $\mathrm{sp}(A)$ is the outcome space of A, the set of possible results of an A-experiment, $\mathcal{M}_A$, a σ-algebra of subsets of A, which are the measurable subsets of $\mathrm{sp}(A)$. For any measurable set $\Delta \in \mathcal{M}_A$, we can define $p^A_\Delta\colon \Omega \to \mathbb{R}$ by

$$p^A_\Delta(\omega) = \alpha_A(\omega)(\Delta).$$

This is an affine mapping yielding the probability, in state ω, that a measurement of A will yield a result in Δ. Since $0 \leq p^A_\Delta(\omega) \leq 1$ for all $\omega \in \Omega$, p^A_Δ is an effect. Functions of an observable are readily definable; for any measurable function f on $\mathrm{sp}(A)$, $f(A)$ is defined to be the observable whose outcome space is $f(\mathrm{sp}(A))$, with the probability that $f(A) \in \Delta$ equal to the probability that $A \in f^{-1}(\Delta)$. A set $\{A_i\}$ of observables is said to be a *compatible* set iff there is an observable C and measurable functions $\{f_i\}$ on $\mathrm{sp}(C)$ such that $A_i = f_i(C)$. Where A is real-valued – that is, $\mathrm{sp}(A) \subseteq \mathbb{R}$ – we will write $\omega(A)$ for the expectation value of A in state ω.

One also specifies a set of affine linear mappings of the state space into itself, representing the physically possible operations on the system. Given a

transformation $T\colon \Omega \to \Omega$, we define the conjugate transformation $T^\dagger\colon E(\Omega) \to E(\Omega)$, which maps effects to effects, by

$$T^\dagger a(\omega) = a(T\omega)$$

for all $\omega \in \Omega$. Obviously, the identity will be among the possible transformations, and it is assumed that transformations can be composed. It is not assumed that all transformations are invertible.

To sum up: with a physical system is associated a triplet $\langle \Omega, \mathcal{O}, \mathcal{T} \rangle$, where Ω is a convex set representing the state space of the system, $\mathcal{O}$ is a set of observables on Ω, and $\mathcal{T}$ is a semi-group of transformations of Ω.

Given two physical systems $S = \langle \Omega, \mathcal{O}, \mathcal{T} \rangle$ and $S' = \langle \Omega', \mathcal{O}', \mathcal{T}' \rangle$, there is a plurality of choices of state space for the composite system. At minimum the set of experiments that can be performed on the combined system should include experiments performed on S and S' separately. Consider the set of experiments in $\mathcal{O}$ with outcome space $\{0, 1\}$ (the set of "yes–no" experiments). For any such experiment A, there is an effect $a \in E(\Omega)$ such that, for any $\omega \in \Omega$, $a(\omega)$ is the expectation value of A in state ω. Since we want to consider theories in which the set of permitted observables falls short of all observables definable on the state space Ω, we want to leave open the possibility that not every effect in $E(\Omega)$ yields the expectation value for some yes–no experiment. Let $\mathcal{E}(\Omega)$ be the set of effects on Ω that *do* figure in this way in some experiments in $\mathcal{O}$. We specify a state in a tensor-product space $\Omega \otimes \Omega'$ by specifying, for every pair $a \in \mathcal{E}(\Omega)$, $b \in \mathcal{E}(\Omega')$, an expectation value for the product observable $a \otimes b$: $a \otimes b$ is measured by measuring a and b separately, and the outcome of the $a \otimes b$ experiment is taken to be the product of the component experiments. A state ω in $\Omega \otimes \Omega'$ is defined to be an affine bilinear mapping $\omega\colon \mathcal{E}(\Omega) \times \mathcal{E}(\Omega') \to [0, 1]$, normalized so that $\omega(u, u') = 1$.[3] Given $\alpha \in \Omega$, $\beta \in \Omega'$, we define the *product state* $\alpha \otimes \beta$ by

$$(\alpha \otimes \beta)(a, b) = \alpha(a)\beta(b)$$

for all $a \in \mathcal{E}(\Omega)$, $b \in \mathcal{E}(\Omega')$. The *minimal tensor product* $\Omega \otimes_{\text{sep}} \Omega'$ is defined to be the convex hull of the set of product states. A state in $\Omega \otimes \Omega'$ is said to be *separable* iff it is in this minimal tensor product; *entangled*, otherwise. The *maximal tensor product* $\Omega \otimes_{\max} \Omega'$ contains all normalized affine bilinear mappings $\omega\colon E(\Omega) \times E(\Omega') \to \mathbb{R}$. In general, a theory may specify a tensor-product space that is a convex proper subset of the maximal tensor-product space. The quantum tensor product, since it includes entangled states, exceeds the minimal tensor product.

[3] Here we depart slightly from Barnum *et al.* [5], who define a tensor-product state as an affine bilinear mapping defined on all pairs of effects on the component spaces.

It falls short of the maximal tensor product, however; see Barnum *et al.* [6] for discussion.

The possible transformations of the composite system should, at minimum, include transformations performed separately on the individual systems. For any transformations $T \in \mathcal{T}$, $T' \in \mathcal{T}'$, define the transformation $T \otimes T'$: $\Omega \otimes_{\max} \Omega' \to \Omega \otimes_{\max} \Omega'$ by

$$(T \otimes T')\omega(a, b) = \omega(T^\dagger a, T'^\dagger b). \tag{9.1}$$

We will require our tensor-product space to be closed under $T \otimes T'$, for all $T \in \mathcal{T}$, $T' \in \mathcal{T}'$.

In his opening lecture at the Boston University conference on Foundations of Quantum Information and Entanglement held in 2006, Abner Shimony raised the question of whether there could be entanglement without potentiality. If by "potentiality" we mean that there are pure states in which some pure observables do not take on definite values, then it becomes a theorem in the convex-set framework that there is no entanglement without potentiality.

Theorem 1. *Let A and B be physical systems, with state spaces Ω_A and Ω_B. If ψ is an entangled pure state in $\Omega_A \otimes \Omega_B$, then there exists a pure effect $a \in E(\Omega_A \otimes \Omega_B)$ such that $0 < \psi(a) < 1$.*

Proof. Let ψ_A and ψ_B be the marginals of ψ, defined by

$$\psi_A(a) = \psi(a, u_B),$$
$$\psi_B(b) = \psi(u_A, b)$$

for all effects $a \in E(\Omega_A)$, $b \in E(\Omega_B)$. Here u_A and u_B are the unit effects in $E(\Omega_A)$ and $E(\Omega_B)$, respectively. By Lemma 3 of Barnum *et al.* [5], if either marginal is pure, then ψ is a product state. Since, by assumption, ψ is entangled, both marginals are mixed.

As ψ_A is a mixed state, let it be a non-trivial mixture of distinct states $\rho, \sigma \in \Omega_A$. If $\rho(a) = \sigma(a)$ for all pure effects a, then $\rho(a) = \sigma(a)$ for *all* effects a, and $\rho = \sigma$. Therefore, there exists a pure effect a such that $\rho(a) \neq \sigma(a)$. It follows from this that $\psi_A(a)$ is not equal to 0 or 1, as it could take on one of these extremal values only if both $\rho(a)$ and $\sigma(a)$ took on the same extremal value. Thus, $\psi(a, u_B) = \psi_A(a) \in (0, 1)$.

The state space of an algebra, be it a C^*-algebra, a JB algebra, or a Segal algebra, is a convex set. Alfsen and Shultz [2] characterize the state spaces of JB algebras among convex sets, and Alfsen, Hanche-Olsen, and Shultz [1] characterize those convex sets that are the state spaces of C^*-algebras. This gives us a way of linking the convex-set approach to the algebraic approach. We will here not go into the full

details of these characterizations, which are presented in Alfsen and Shultz [3], but focus instead on a central feature of these characterizations.

First, some definitions. A *face* of a convex set Ω is a convex subset $F \subseteq \Omega$ that is such that, if any state in F is a mixture of states $\rho, \omega \in \Omega$, then ρ and ω are also in F. The face generated by ρ, ω, is the smallest face containing ρ and ω. In a simplex, the face generated by two pure states is just the one-dimensional simplex consisting of ρ, ω, and all mixtures of the two. The face generated by two pure quantum states, represented by state vectors $|\psi\rangle$ and $|\phi\rangle$, consists of a set of pure states comprising all linear superpositions of $|\psi\rangle$ and $|\phi\rangle$, together with all mixtures of these. Thus, it is affinely isomorphic to the Bloch sphere. A face $F \subseteq \Omega$ is *norm exposed* iff there is an effect $a \in E(\Omega)$ such that F is precisely the set of states on which a takes on the value 0. A *Hilbert ball* is a convex set of states that is affinely isomorphic to the closed unit ball of some (finite- or infinite-dimensional) real Hilbert space. The Hilbert ball B^n is the closed unit ball in $\mathbb{R}^n$.

If Ω is the state space of a JB algebra, or, as a special case, a C^*-algebra, then the faces generated by pairs of pure states take on a simple form.

Theorem 2. *If a convex set Ω is the state space of a JB algebra, then for any distinct pure states $\rho, \omega \in \Omega$, the face generated by ρ, ω is a norm-exposed Hilbert ball.*

For proof, see Alfsen *et al.* [2], or Alfsen and Shultz ([3], Proposition 9.10). For discussion of the significance of this result for the Spekkens and Schrödinger theories, see Halvorson [16].

Theorem 3. *If a convex set Ω is the state space of a C^*-algebra, then, for any distinct pure states $\rho, \omega \in \Omega$, the face generated by ρ, ω is norm exposed, and is either B^1 or B^3.*

See Alfsen *et al.* [1], or Alfsen and Shultz [3], Theorem 11.59.

Theorem 3 entails that, in the state space of a C^*-algebra, the face generated by a pair of pure states ρ and ω is either the one-dimensional simplex consisting of ρ and ω and all mixtures of these, or else is affinely isomorphic to the Bloch sphere. Thus, within the C^*-algebra framework, the state space of quantum mechanics can be completely characterized by the condition that any two pure states ρ and σ are connected by a continuous path through the set of pure states within the face generated by ρ and σ – a cousin of Hardy's [19] continuity axiom. For a classical theory, on the other hand, the set of pure states is totally disconnected: any state that is "intermediate" between two classical states is a mixture of the two.

Consider a system S, with state space Ω_S, and let A be some subsystem of S. The state space Ω_S will be $\Omega_A \otimes \Omega_E$, where E is the "environment" of A, that is, everything in S that is not included in the subsystem A. For any pure state $\varepsilon \in \Omega_E$, the set $\Omega_A \otimes \varepsilon$ is a face of Ω_S that is affinely isomorphic to Ω_A. Thus, the state space of any subsystem of S will be affinely isomorphic to a face of Ω_S. The simplest subsystems will be those whose state spaces are the faces generated by a pair of distinguishable states. Theorem 3 therefore says that, in the state space of a C^*-algebra, these simplest systems will have state spaces that are equivalent either to the state space of a classical bit, or to that of a qubit. Composite systems might be hybrids of the two.

This is another reason for regarding the C^*-algebraic framework as too restrictive for a framework meant to embrace physical theories in general. This limitation is also a strength. Though it can make no claims to physical generality, the C^*-framework is a useful one for comparing classical and quantum theories, and for discussing information processing and transmission possibilities afforded by a combination of classical and quantum operations, precisely because it is a framework in which the simplest systems are constrained to be either classical bits or qubits.

The work of Clifton, Bub, and Halvorson [11] was motivated by a suggestion, which they attribute to Chris Fuchs and Gilles Brassard, that quantum mechanics can be derived from an appropriately chosen set of cryptographic principles. The C^*-algebraic framework is too restrictive a starting point for such a project. A more appropriate starting point is the convex-set approach. The following question arises: is there a set of principles, with clear operational (cryptographic or otherwise) significance, that distinguishes quantum state spaces among the convex sets?

Alfsen and Shultz [2] and Araki [4] tell us how to single out, among convex sets, those that are the state spaces of JB algebras. Alfsen *et al.* [1] tell us how to distinguish, among these, those that are state spaces of C^*-algebras. The further condition that the face generated by any pair of pure states is affinely isomorphic to the Bloch sphere distinguishes the quantum state spaces among these. These theorems, therefore, could form a useful starting point for a project that seeks to characterize quantum mechanics in terms of information-theoretic principles. Araki's work may be useful in this regard because his conditions are expressed in terms more directly connected with operations than those of Alfsen *et al.* The information-theoretic principles proposed by Clifton, Bub, and Halvorson do not suffice to characterize quantum mechanics, because, even within the C^*-framework, they do not entail that there are *no* states that are dispersion-free in all observables. The task of characterizing quantum mechanics in terms of possibilities of information processing and transfer remains to be achieved.

9.5 The Spekkens toy theory

Consider a ball that can be in one of four boxes. A state of this system is specified by four probabilities (p_1, p_2, p_3, p_4). Extremal states are those in which the ball is definitely in one of the boxes. The state space is therefore a simplex consisting of four pure states and mixtures of these, and can be depicted as a tetrahedron.

Now, suppose we impose a restriction on state preparations, to the effect that the probability of the ball being in any one of the boxes cannot exceed $\frac{1}{2}$. This restriction chops the vertices off our tetrahedron, leaving an octahedron whose vertices are the midpoints of each edge of our original tetrahedral state space (see Figure 9.1). These six states, which are the extremal points of the new state space, are those in which the probability is equally divided between two of the boxes. These are the pure states of elementary systems in the toy theory constructed by Spekkens [29, 30], the "states of maximal knowledge," as Spekkens puts it. In Spekkens' theory, arbitrary mixtures of these pure states are not permitted; Spekkens permits as a mixture only the state in which the probability is equally divided among all four boxes. This seems to be an inessential restriction. Moreover, it is one of questionable coherence, if the states are interpreted, as Spekkens would have them be, as possible states of knowledge. A theory whose states are intended as states of knowledge ought to include states of less-than-maximal knowledge. Suppose Alice's belief state is a Spekkens maximal-knowledge state about an elementary system of Spekkens' theory, that Bob knows nothing about the system except that Alice's belief state is either ρ_1 or ρ_2, and that Bob's degree of belief in the former is λ, in the latter $1 - \lambda$. Then Bob's epistemic state about the system is the corresponding mixture of states of maximal knowledge.

We will, therefore, take as a state space the convex hull of Spekkens' pure states, as suggested by Halvorson [16]. This is the octahedron we have just described, depicted again in Figure 9.2. Note that, though we started with a simplex, the state space we have ended up with is *not* a simplex. In particular, the maximally mixed

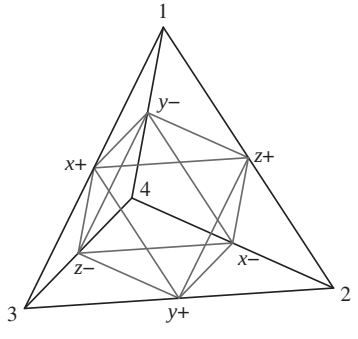


Fig. 9.1 The state space of the Spekkens toy theory.

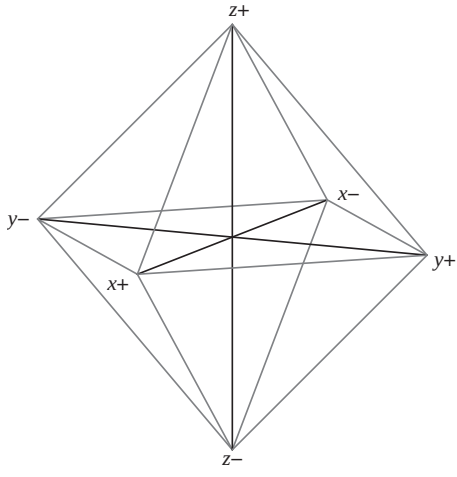


Fig. 9.2 The octahedron reoriented.

state in the center (call it ω_0), can be decomposed as an equally weighted mixture of $\{x+, x-\}$, or of $\{y+, y-\}$, or of $\{z+, z-\}$.

Given this state space, one can define observables that do not extend to observables on our original tetrahedral simplex, and hence do not correspond to observable quantities in the Spekkens theory. There is, for example, an affine functional on the octahedron that takes on the value 0 on the $(x+, y+, z+)$ face and the value 1 on the $(x-, y-, z-)$ face. Were such a function extended to the embedding tetrahedron, it would have to take on a value greater than 1 on vertex 4. The set of permitted observables will be a proper set of the observables definable on the state space. Define the affine functionals[4]

$$\begin{aligned} &a_x(x+) = 1, \qquad a_x(x-) = 0, \qquad a_x(y+) = a_x(z+) = \frac{1}{2}, \\ &a_y(y+) = 1, \qquad a_y(y-) = 0, \qquad a_y(x+) = a_y(z+) = \frac{1}{2}, \\ &a_z(z+) = 1, \qquad a_z(z-) = 0, \qquad a_z(x+) = a_z(y+) = \frac{1}{2}. \end{aligned} \tag{9.2}$$

We define the observables $\{\sigma_\mu | \mu = x, y, z\}$ analogous to the quantum spin observables. The outcome space of σ_μ is $\{-1, +1\}$, and the probability in state ω of obtaining outcome $+1$ upon measuring σ_μ is $a_\mu(\omega)$. Take as our set of observables these three, together with the identity observable I, which yields with certainty the result $+1$ in any state.

[4] The affine dimension of the Spekkens state space, like that of the Bloch sphere, is four. That is, to specify any affine function, it suffices to specify its value on a set of four linearly independent states. We could have chosen the same set of four for the definition of each of these functionals, say, $\{x+, y+, z+, \omega_0\}$.

Now suppose we start with the Bloch sphere, which represents the state space of a single qubit (equivalently, a spin-$\frac{1}{2}$ particle), and impose a restriction that the only preparable pure states are spin eigenstates in one of three mutually orthogonal directions: our state-preparation device is a Stern–Gerlach apparatus whose mounting permits only these three orientations. Call these directions (x, y, z). Label the six preparable pure states $\{x+, x-, y+, y-, z+, z-\}$. The convex hull of these states is once again our octahedron. We place the same restrictions on measurement: only spin measurements in the same three directions are permitted. What we thereby obtain is a state space and set of observables that are isomorphic to those of elementary systems of the Spekkens theory. Thus, the same structure, consisting of state space plus set of permitted observables, is obtainable on the one hand by starting from a simplex, that is, an essentially classical state space, and imposing restrictions on state preparation, and on the other by starting from a quantum state space and imposing restrictions on state preparation and measurement. Moreover, as Spekkens [29] demonstrates in detail, many of the features of quantum theories that one might be tempted to think of as characteristic of quantum mechanics already exist within this reduced state space.

Though we have obtained the same structure of state space and observables in the two cases, the resulting theories will nevertheless be different if the permitted transformations of the state spaces are different, and, since we require tensor-product spaces to be closed under transformations of the component systems, this will have consequences for the state spaces of composite systems. Unitary transformations of the quantum state space of a qubit correspond to rotations of the Bloch sphere. The transformations permitted in the Spekkens theory are permutations of the four boxes. We thus have two natural choices for the group of permitted transformations of our octahedral state space: $\mathcal{T}_S$, the group of transformations resulting from permutations of the four boxes, or $\mathcal{T}_q$, the set of rotations that take the octahedron into itself. It turns out that these are distinct representations of the same group.[5]

Obviously, any automorphism of the state space will have to leave the unique totally mixed state ω_0 fixed. To specify an automorphism, therefore, it suffices to specify its action on the set of pure states $\{x+, y+, z+\}$, which any automorphism will take into a set of pure states containing no pair of opposite states. Both groups of transformations include cyclic permutations of $x+, y+, z+$. Let T_1 be the transformation $\langle x+ \to y+, y+ \to z+, z+ \to x+\rangle$. This corresponds, in the quantum domain, to a rotation of $2\pi/3$ about the axis pointing in the direction of $\hat{x} + \hat{y} + \hat{z}$. It is achieved by the Spekkens transformation that leaves box 4 invariant and permutes the other three by $\langle 1 \to 3 \to 2 \to 1\rangle$. Let T_2 be the transformation that takes $x+$ to $x-$ and $y+$ to $y-$, and leaves $z+$ invariant. This is achieved, in the group

[5] I would like to thank Rob Spekkens for pointing this out.

Table 9.1 *Generators of $\mathcal{T}_S$ and $\mathcal{T}_q$*

	T			$T^\dagger$		
	$x+$	$y+$	$z+$	σ_x	σ_y	σ_z
T_1	$y+$	$z+$	$x+$	σ_z	σ_x	σ_y
T_2	$x-$	$y-$	$z+$	$-\sigma_x$	$-\sigma_y$	σ_z
T_3^S	$y+$	$x-$	$z-$	$-\sigma_y$	σ_x	$-\sigma_z$
T_3^q	$y+$	$x-$	$z+$	$-\sigma_y$	σ_x	σ_z

of rotations, by a rotation of π about the z-axis, and, in the group of Spekkens permutations, by the swap $\langle 1 \leftrightarrow 2, 3 \leftrightarrow 4\rangle$. Rotations about the other axes can be achieved by combining this rotation with cyclic permutations of the axes; we thereby achieve any transformation that inverts two axes while leaving the other invariant.

Let T_3^q be a counterclockwise rotation of $\pi/2$ about the z-axis. This takes $x+$ to $y+$ and $y+$ to $x-$, and leaves $z+$ invariant. It is not achievable via any permutation of boxes, and so does not belong to the Spekkens group of transformations. We do, however, have in the Spekkens group a transformation that consists of a $\pi/2$ rotation about the z-axis followed by a reflection in the xy plane; this is achieved by the permutation $1 \to 3 \to 2 \to 4 \to 1$. Call this T_3^S. Note that T_3^q and T_3^S, applied twice, both yield T_2.

The reader will be able to verify that the sets $\{T_1, T_3^S\}$ and $\{T_1, T_3^q\}$ suffice to generate the Spekkens and quantum groups of transformations, respectively. We summarize the generating sets of the Spekkens and quantum transformations in Table 9.1, in which, for each automorphism T of the state space Ω, we include also the conjugate automorphism $T^\dagger$.

Let us now begin to construct tensor-product spaces for the two theories. A state is specified by specfying its value on the effects $a_\mu \otimes a_\nu$, $\mu, \nu \in \{0, x, y, z\}$ (where we take a_0 to be the unit effect u), or, equivalently, by specifying the expectation value it assigns to the observables $\sigma_\mu \otimes \sigma_\nu$, $\mu, \nu \in \{0, x, y, z\}$, where σ_0 is the observable yielding outcome 1 in all states. The minimal tensor product contains all product states and convex combinations of product states. In addition to these, Spekkens allows states whose marginals are maximally mixed, but with maximal correlations between the states of the component systems. One such state is the state (call it χ) in which it is certain that both balls are in boxes with the same label (that is, if ball 1 is in its box 1, then ball 2 is in its box 1 also, etc.), with probabilities equally divided among the four possibilities. This state is symbolized in Figure 9.3. In this 4×4 array, system 1's box states are along the vertical axis,

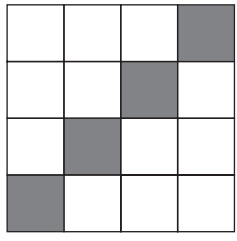


Fig. 9.3 The maximally correlated state χ.

and system 2's box states along the horizontal; for the joint system to occupy the box (i, j) thus represents a state in which ball 1 is in box j and ball 2 is in box i. The state χ is defined by

$$\chi(\sigma_\mu \otimes \sigma_\nu) = \delta_{\mu\nu}, \quad \mu, \nu \in \{0, x, y, z\}. \tag{9.3}$$

That is, the three observables $\{\sigma_x \otimes \sigma_x, \sigma_y \otimes \sigma_y, \sigma_z \otimes \sigma_z\}$ take on dispersion-free values, equal to 1. Note that, since, quantum-mechanically,

$$(\sigma_x \otimes \sigma_x)\big(\sigma_y \otimes \sigma_y\big) = -\sigma_z \otimes \sigma_z, \tag{9.4}$$

this is not a state in the standard quantum tensor product.

If we include the state χ in our tensor-product space, then all states obtainable from it by permutations of the boxes will also be in the tensor-product space (we can restrict our attention to permutations on one subsystem, since we will not obtain any new states by considering permutations of both subsystems). This gives us a set of 24 pure, maximally entangled states, each of which can be represented by a 4×4 array with one element of each row and each column shaded. For example, since we have the state χ, application of T_2 produces the state in which $\sigma_x \otimes \sigma_x$ and $\sigma_y \otimes \sigma_y$ have definite value -1, and $\sigma_z \otimes \sigma_z$ has definite value $+1$, and, by combining T_1 with T_2, we obtain common eigenstates of these observables for other combinations of eigenvalues that multiply to $+1$ (note that, since the only sign-changing element of our generating set for the Spekkens group is T_3^S, and this changes two signs, we will not be able to change the sign of the product of the three eigenvalues). These four states are depicted in Figure 9.4.

By application of T_3^S to χ, we get the $- + -$ common eigenstate of $\{\sigma_y \otimes \sigma_x, \sigma_x \otimes \sigma_y, \sigma_z \otimes \sigma_z\}$, and via applications of T_1 and T_2, the other eigenstates with eigenvalues that multiply to $+1$. These states are depicted in Figure 9.5. By continuing the process, we obtain, for each of the rows and each of the columns of

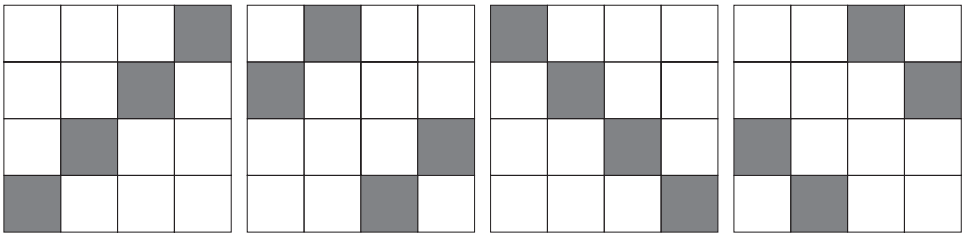


Fig. 9.4 Common eigenstates of $\sigma_x \otimes \sigma_x$, $\sigma_y \otimes \sigma_y$, $\sigma_z \otimes \sigma_z$.

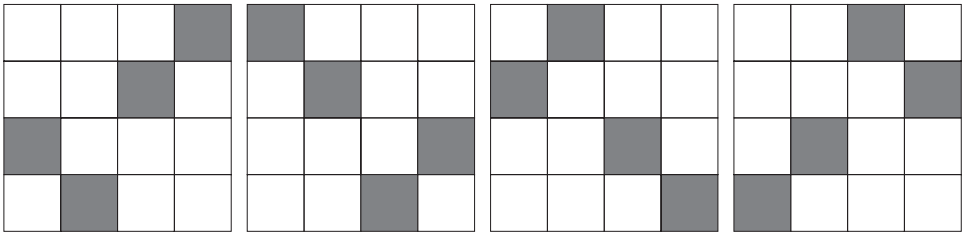


Fig. 9.5 Common eigenstates of $\sigma_x \otimes \sigma_y$, $\sigma_y \otimes \sigma_x$, $\sigma_z \otimes \sigma_z$.

the following array of observables, four common eigenstates of the observables in that row (column), corresponding to each combination of eigenvalues multiplying to $+1$,

$$\begin{array}{lll} \sigma_x \otimes \sigma_x & \sigma_y \otimes \sigma_y & \sigma_z \otimes \sigma_z \\ \sigma_z \otimes \sigma_y & \sigma_x \otimes \sigma_z & \sigma_y \otimes \sigma_x \\ \sigma_y \otimes \sigma_z & \sigma_z \otimes \sigma_x & \sigma_x \otimes \sigma_y \end{array} \tag{9.5}$$

These states are the 24 pure entangled states of the Spekkens theory. Call the tensor-product space whose pure states consist of the product states together these 24 maximally entangled states, Ψ_S. Note that the rows and columns of (9.5) are, in quantum theory, also sets of observables having joint eigenstates. In the quantum theory, however, eigenvalues of observables in a row will multiply, not to $+1$, but to -1.

Suppose that we begin constructing a quantum-like tensor-product space for our theory by starting with a pair of systems having our octahedral state space, and endowing it with a state that *is* in the quantum tensor product. For example, we

could start with the state ξ, common to both the quantum and the Spekkens tensor product, defined by

$$\begin{aligned} &\xi(\sigma_x \otimes \sigma_y) = \xi(\sigma_y \otimes \sigma_x) = \xi(\sigma_z \otimes \sigma_z) = +1, \\ &\xi(\sigma_\mu \otimes \sigma_\nu) = 0, \text{ all other combinations.} \end{aligned} \tag{9.6}$$

Application of the quantum group of transformations again yields common eigenstates for each of the rows and columns of the array (9.5). For example, applying T_3^q to ξ yields the state ϕ^-, with

$$\phi^-(\sigma_x \otimes \sigma_x) = -1, \qquad \phi^-(\sigma_y \otimes \sigma_y) = +1, \qquad \phi^-(\sigma_z \otimes \sigma_z) = +1.$$

Pairwise sign changes (combining applications of T_1 and T_2) yield the other common eigenstates of $\{\sigma_x \otimes \sigma_x, \sigma_y \otimes \sigma_y, \sigma_z \otimes \sigma_z\}$, forming the set of Bell states $\{\phi^+, \phi^-, \psi^+, \psi^-\}$, which are defined by the conditions

$$\begin{aligned} &\phi^+(\sigma_x \otimes \sigma_x) = +1, \quad \phi^+(\sigma_y \otimes \sigma_y) = -1, \quad \phi^+(\sigma_z \otimes \sigma_z) = +1, \\ &\phi^-(\sigma_x \otimes \sigma_x) = -1, \quad \phi^-(\sigma_y \otimes \sigma_y) = +1, \quad \phi^-(\sigma_z \otimes \sigma_z) = +1, \\ &\psi^+(\sigma_x \otimes \sigma_x) = +1, \quad \psi^+(\sigma_y \otimes \sigma_y) = +1, \quad \psi^+(\sigma_z \otimes \sigma_z) = -1, \\ &\psi^-(\sigma_x \otimes \sigma_x) = -1, \quad \psi^-(\sigma_y \otimes \sigma_y) = -1, \quad \psi^-(\sigma_z \otimes \sigma_z) = -1, \end{aligned} \tag{9.7}$$

together with the condition that, for $\mu \neq \nu$, $\phi^+(\sigma_\mu \otimes \sigma_\nu) = \phi^-(\sigma_\mu \otimes \sigma_\nu) = \psi^+(\sigma_\mu \otimes \sigma_\nu) = \phi^-(\sigma_\mu \otimes \sigma_\nu) = 0$. Let Ψ_q be the tensor-product space whose set of pure entangled states is the closure of the set of Bell states under $\mathcal{T}_q \otimes I$. Like the Spekkens tensor product, Ψ_q contains 24 pure entangled states, consisting of four common eigenstates for each of the rows and columns of (9.5). It shares with Ψ_S the eigenstates of observables in the columns of the array; it differs from Ψ_S in having eigenvalues of observables in any row of (9.5) that multiply to -1.

An alternative tensor product (call it $\tilde{\Psi}_q$) can be constructed by including the state χ and taking the closure under the quantum set of transformations. In this tensor product, joint eigenstates of the rows of (9.5) will have eigenvalues multiplying to $+1$, with eigenvalues of the columns multiplying to -1; this set of entangled states consists of those obtainable from the quantum states by a parity inversion on one of the component systems. Choosing one or the other of these tensor products amounts to a choice of relative orientation on the two systems – that is, given an orientation of one, a choice of which orientation of the other will be regarded as the "same" orientation. We can also construct a tensor product by including *both* sets of entangled states.

Similar remarks apply to tensor products closed under the Spekkens transformations: there is one such space, Ψ_S, containing χ and eigenstates of all rows and columns of (9.5) with sets of eigenvalues multiplying to $+1$; another, $\tilde{\Psi}_S$, containing the Bell states and eigenstates of all rows and columns of (9.5) with eigenvalues multiplying to -1; and one that is the convex hull of $\Psi_S \cup \tilde{\Psi}_S$.

Which tensor product we use will have consequences for whether or not a Kochen–Specker obstruction will be forthcoming – it will be possible to construct a Kochen–Specker obstruction in the spaces Ψ_q and $\tilde{\Psi}_q$, but not in Ψ_S or $\tilde{\Psi}_S$.

Mermin's simple Kochen–Specker obstruction [25], which is applicable to the Hilbert space of a pair of qubits, begins with the observation that, in the following array, each row and each column consists of mutually compatible quantum observables:

$$\begin{array}{ccc} \sigma_x \otimes I & I \otimes \sigma_x & \sigma_x \otimes \sigma_x \\ I \otimes \sigma_y & \sigma_y \otimes I & \sigma_y \otimes \sigma_y \\ \sigma_x \otimes \sigma_y & \sigma_y \otimes \sigma_x & \sigma_z \otimes \sigma_z \end{array} \tag{9.8}$$

Because of the algebraic relations

$$\sigma_x \sigma_y = -\sigma_y \sigma_x = \mathrm{i}\sigma_z, \tag{9.9}$$

we have

$$\begin{array}{l} (\sigma_x \otimes \sigma_x)(\sigma_y \otimes \sigma_y) = -\sigma_z \otimes \sigma_z, \\ (\sigma_x \otimes \sigma_y)(\sigma_y \otimes \sigma_x) = \sigma_z \otimes \sigma_z. \end{array} \tag{9.10}$$

The product of the elements of each row of the array (9.8) is the identity $I \otimes I$, as is the product of the elements of each of the first two columns, whereas the product of the elements of the third column is $-I \otimes I$. This means that, if we try to assign definite values to these nine observables, satisfying the product rule that, whenever A and B are compatible observables, $v(AB) = v(A)v(B)$, we cannot succeed; the product of all nine of these values would have to be both $+1$ and -1.

Return now to the convex-set framework and the tensor products we have constructed for pairs of octahedral state spaces. We have not introduced any notion of multiplication of observables, and it is not clear how to make sense, in the general setting, of multiplication of incompatible observables. Algebraic relations among compatible observables, however, *do* make sense. Recall that a set $\{A_i\}$ of observables is a compatible set if and only if there is an observable C such that each A_i is a function of C. This will give rise to functional relations among the observables A_i.

Consider, now, the four Bell states $\{\phi^+, \phi^-, \psi^+, \psi^-\}$. For each of these states, there is an affine functional that takes on the value 1 on that state and 0 on the others: define $\{a_{\phi^+}, a_{\phi^-}, a_{\psi^+}, a_{\psi^-}\}$ by

$$\begin{aligned}
a_{\phi^+}(\omega) &= \tfrac{1}{4}\big(1 + \omega\,(\sigma_x \otimes \sigma_x) - \omega\,\big(\sigma_y \otimes \sigma_y\big) + \omega\,(\sigma_z \otimes \sigma_z)\big),\\
a_{\phi^-}(\omega) &= \tfrac{1}{4}\big(1 - \omega\,(\sigma_x \otimes \sigma_x) + \omega\,\big(\sigma_y \otimes \sigma_y\big) + \omega\,(\sigma_z \otimes \sigma_z)\big),\\
a_{\psi^+}(\omega) &= \tfrac{1}{4}\big(1 + \omega\,(\sigma_x \otimes \sigma_x) + \omega\,\big(\sigma_y \otimes \sigma_y\big) - \omega\,(\sigma_z \otimes \sigma_z)\big),\\
a_{\psi^-}(\omega) &= \tfrac{1}{4}\big(1 - \omega\,(\sigma_x \otimes \sigma_x) - \omega\,\big(\sigma_y \otimes \sigma_y\big) - \omega\,(\sigma_z \otimes \sigma_z)\big).
\end{aligned} \tag{9.11}$$

These four affine functionals sum to the unit functional,

$$a_{\phi^+} + a_{\phi^-} + a_{\psi^+} + a_{\psi^-} = u. \tag{9.12}$$

If they are effects – that is, if their values on *any* state are confined to the interval $[0, 1]$ – then we will be able to define a discrete observable C with outcome space $\mathrm{sp}(C) = \{1, 2, 3, 4\}$, with probabilities for the four outcomes yielded by the four effects $\{a_{\phi^+}, a_{\phi^-}, a_{\psi^+}, a_{\psi^-}\}$. If there is such an observable C, the Bell states are a distinguishable set of states. The observables $\{\sigma_x \otimes \sigma_x, \sigma_y \otimes \sigma_y, \sigma_z \otimes \sigma_z\}$ will be functions of C: take f_i to be the functions on $\mathrm{sp}(C)$, defined by

$$\begin{aligned}
f_1(n) &= 2(\delta_{1n} + \delta_{3n}) - 1,\\
f_2(n) &= 2(\delta_{2n} + \delta_{3n}) - 1,\\
f_3(n) &= 2(\delta_{1n} + \delta_{2n}) - 1.
\end{aligned} \tag{9.13}$$

Then we will have

$$\begin{aligned}
\sigma_x \otimes \sigma_x &= f_1(C),\\
\sigma_y \otimes \sigma_y &= f_2(C),\\
\sigma_z \otimes \sigma_z &= f_3(C).
\end{aligned} \tag{9.14}$$

Moreover, since

$$f_1(x)\,f_2(x) = -f_3(x), \tag{9.15}$$

for all $x \in \mathrm{sp}(C)$, we will have the functional relation

$$f_1(C)\,f_2(C) = -f_3(C), \tag{9.16}$$

or

$$(\sigma_x \otimes \sigma_x)\,\big(\sigma_y \otimes \sigma_y\big) = -\sigma_z \otimes \sigma_z. \tag{9.17}$$

It is easy to check that, on the tensor product Ψ_q, the functionals $\{a_{\phi^+}, a_{\phi^-}, a_{\psi^+}, a_{\psi^-}\}$ are indeed effects. Therefore, we can add to the observables on Ψ_q an observable C that distinguishes the Bell states, and, having done so, we will obtain the algebraic relations (9.17). In the tensor product $\tilde{\Psi}_q$, each of these functionals takes on the value $-1/2$ on some state, so they are not effects. We will, however, be able to define effects on $\tilde{\Psi}_q$ that lead to the "anti-quantum" functional relation $(\sigma_x \otimes \sigma_x)\left(\sigma_y \otimes \sigma_y\right) = \sigma_z \otimes \sigma_z$.

In a similar manner, we can define affine functionals that distinguish Ψ_q's common eigenstates of $\{\sigma_x \otimes \sigma_y, \sigma_y \otimes \sigma_x, \sigma_z \otimes \sigma_z\}$ and, on Ψ_q, are effects, and use these to add an observable that induces the functional relation

$$\left(\sigma_x \otimes \sigma_y\right)\left(\sigma_y \otimes \sigma_x\right) = \sigma_z \otimes \sigma_z. \tag{9.18}$$

It is worth noting in passing that, if we move to the larger tensor-product space that is the convex hull of $\Psi_q \cup \tilde{\Psi}_q$, it will no longer be possible to introduce an observable that distinguishes the Bell states. This can be seen from the fact that any affine functional that takes on the value 1 on ϕ^+ and 0 on the other Bell states will take on the value $-1/2$ on the state $\tilde{\phi}^+$, defined by

$$\begin{aligned} \tilde{\phi}^+(I \otimes \sigma_i) &= \tilde{\phi}^+(\sigma_j \otimes I) = 0 \\ \tilde{\phi}^+(\sigma_i \otimes \sigma_j) &= -\phi^+(\sigma_i \otimes \sigma_j). \end{aligned} \tag{9.19}$$

for $i, j \in \{x, y, z\}$. This follows from the fact that the totally mixed state $\omega_0 \otimes \omega_0$ can be written either as an equally weighted mixture of the Bell states, or as an equally weighted mixture of ϕ^+ and $\tilde{\phi}^+$:

$$\omega_0 \otimes \omega_0 = \frac{1}{4}\left(\phi^+ + \phi^- + \psi^+ + \psi^-\right) = \frac{1}{2}\left(\phi^+ + \tilde{\phi}^+\right). \tag{9.20}$$

Similar remarks hold for the convex hull of $\Psi_S \cup \tilde{\Psi}_S$. On this space, there can be no positive affine functional that takes on the value 1 on the state ξ and 0 on all the other eigenstates of $\{\sigma_x \otimes \sigma_x, \sigma_y \otimes \sigma_y, \sigma_z \otimes \sigma_z\}$.

By adding to the set of observables on Ψ_q an observable that distinguishes the Bell states, and one that distinguishes the common eigenstates of $\{\sigma_x \otimes \sigma_y, \sigma_y \otimes \sigma_x, \sigma_z \otimes \sigma_z\}$, we induce the algebraic relations (9.17) and (9.18), which lead to a Mermin-style Kochen–Specker obstruction. It is worth noting that we have obtained these relations, not by reference to the quantum relation (9.9), which involves products of incompatible observables, but by considerations of relations between *compatible* observables.

In a similar manner, we can induce algebraic relations among the same observables in the Spekkens tensor product Ψ_S. We will, in this space, have the relation (9.18). However, instead of the quantum relation (9.17), we will have

$$(\sigma_x \otimes \sigma_x)\big(\sigma_y \otimes \sigma_y\big) = \sigma_z \otimes \sigma_z, \tag{9.21}$$

and no Kochen–Specker obstruction will be forthcoming.

Thus, starting with the same state space for elementary systems, we obtain, depending on which tensor-product space we bestow on pairs of systems, either a theory whose pure states are those of the Spekkens toy theory, which admit of representations as mixtures of classical states, or a theory, obtained by using one of the quantum tensor products, that admits of no non-contextual hidden-variables theory. Similarly, it is clear that whether or not our theory will contain states with correlations that violate a Bell inequality will depend on the tensor product used – though the tensor products we have so far constructed contain no Bell-inequality-violating correlations, we can, without contradiction (and without disrupting the algebraic relations between observables we have obtained) extend our tensor space (be it Spekkens or quantum) to include them.

All this begins to suggest that Schrödinger was right to find the essential distinction between classical and quantum theory in the way that the latter treats compound systems. However, it is not entanglement *per se* that is distinctively quantum; it matters *which* entangled states the theory contains.

9.6 Appendix

9.6.1 C- and JB algebras*

A *Banach space* is a normed linear vector space that is complete with respect to the norm. That is, every Cauchy sequence converges to a limit.

A *Banach algebra* $\mathfrak{A}$ is a Banach space that is also an algebra with identity I, such that the operation of multiplication is separately continuous. That is, for each $B \in \mathfrak{A}$, if $A_n \to A$, then $A_n B \to AB$, and, for each $A \in \mathfrak{A}$, if $B_n \to B$, then $AB_n \to AB$.

An *involution* is a mapping $A \to A^*$ such that

(i) $(aA + bB)^* = \bar{a}A^* + \bar{b}B^*$
(ii) $(AB)^* = B^*A^*$
(iii) $(A^*)^* = A$

A *C^*-algebra* is a complex Banach algebra with an involution that satisfies $\|A^*A\| = \|A\|^2$.

A *Jordan algebra* is a vector space with a commutative bilinear product $\circ$ satisfying

$$(a^2 \circ b) \circ a = a^2 \circ (b \circ a).$$

A *JB algebra* is a Jordan algebra equipped with a norm $\| \ \|$ satisfying

(i) $\|a \circ b\| \leq \|a\|b\|$
(ii) $\|a^2\| = \|a\|^2$
(iii) $\|a^2\| \leq \|a^2 + b^2\|$,

which is complete with respect to this norm.

The self-adjoint elements of any C^*-algebra form a JB algebra, with the symmetric product

$$a \circ b = \frac{1}{2}(ab + ba).$$

9.6.2 C-algebraic no-cloning*

Theorem 4. *If $\{\rho, \omega\}$ are a clonable pair of distinct pure states of a C^*-algebra then they are orthogonal.*

Proof. We define the *transition probability* of a pair of pure states by

$$p(\rho, \omega) = 1 - \frac{1}{4}\|\rho - \omega\|^2.$$

A non-selective operation cannot increase the norm distance between two states; therefore, the transition probability cannot decrease under a non-selective operation. Hence, if there is a non-selective operation that clones $\{\rho, \omega\}$,

$$p(\rho \otimes \rho, \omega \otimes \omega) \geq p(\rho, \omega).$$

However, it follows from CBH's Lemma 2 that, for pure states ρ, ω,

$$p(\rho \otimes \rho, \omega \otimes \omega) = p(\rho, \omega)^2.$$

This gives us

$$p(\rho, \omega)^2 \geq p(\rho, \omega),$$

or

$$p(\rho, \omega)(1 - p(\rho, \omega)) \leq 0.$$

Since $p(\rho, \omega)$ lies with in the interval [0, 1], this is possible only if $p(\rho, \omega)$ is equal to 0 or 1.

Acknowledgments

I would like to thank Jeremy Butterfield and Rob Spekkens for their comments on an earlier draft, and Howard Barnum, Alex Wilce, Jonathan Barrett, and Matt Leifer for answering many questions about convex sets.

References

[1] E. M. Alfsen, H. Hanche-Olsen, and F. W. Shultz, State spaces of C^*-algebras, *Acta Math.* **144**, 267–305 (1980).

[2] E. M. Alfsen and F. W. Shultz, State spaces of Jordan algebras, *Acta Math.* **144**, 267–305 (1978).

[3] E. M. Alfsen and F. W. Shultz, *Geometry of State Spaces of Operator Algebras*, (Boston, MA: Birkhäuser, 2003).

[4] H. Araki, On a characterization of the state space of quantum mechanics, *Commun. Math. Phys.* **75**, 1–24 (1980).

[5] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, Cloning and broadcasting in generic probabilistic models, quant-ph/0611295 (2006).

[6] H. Barnum, C. A. Fuchs, J. M. Renes, and A. Wilce, Influence-free states on compound quantum systems, quant-ph/0507108 (2005).

[7] J. S. Bell, On the Einstein–Podolsky–Rosen paradox, *Physics* **1**, 195–200 (1964). (Reprinted in [9], pp. 14–21.)

[8] J. S. Bell, On the problem of hidden variables in quantum mechanics, *Rev. Mod. Phys.* **38**, 447–52 (1966). (Reprinted in [9], pp. 1–13).

[9] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge: Cambridge University Press, 1987).

[10] E. G. Beltrametti, and S. Bugajski, Effect algebras and statistical physical theories, *J. Math. Phys.* **38**, 3020–3030 (1997).

[11] R. Clifton, J. Bub, and H. Halvorson, Characterizing quantum theory in terms of information-theoretic constraints, *Foundations Phys.* **33**, 1561–1590 (2003).

[12] G. M. D'Ariano, Operational axioms for quantum mechanics, quant-ph/0611094 (2006).

[13] C. A. Fuchs, Quantum mechanics as quantum information (and only a little more), quant-ph/0205039 (2002).

[14] C. A. Fuchs and R. Schack, Unknown quantum states and operations, a Bayesian view, quant-ph/0404156 (2004).

[15] D. M. Greenberger, M. Horne, and A. Zeilinger, Going beyond Bell's theorem, in *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, M. Kafatos (ed.) (Dordrecht: Kluwer, 1989), pp. 73–76.

[16] H. Halvorson, On information-theoretic characterizations of physical theories, *Stud. Hist. Philos. Mod. Phys.* **35**, 277–293 (2004).

[17] L. Hardy, Quantum mechanics, local realistic theories, and Lorentz-invariant realistic theories, *Phys. Rev. Lett.* **68**, 2981–2984 (1992).

[18] L. Hardy, Nonlocality for two particles without inequalities for almost all entangled states, *Phys. Rev. Lett.* **71**, 1665–1668 (1993).

[19] L. Hardy, Quantum theory from five reasonable axioms, quant-ph/0101012 (2001).

[20] L. Hardy, Why quantum theory?, in *Non-Locality and Modality: Proceedings of the NATO Advanced Research Workshop on Modality, Probability, and Bell's Theorems*, T. Placek and J. Butterfield (eds.) (Berlin: Springer, 2002), pp. 61–74. Also quant-ph/0111068 (2002).

[21] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland, 2002).

[22] D. Lewis, A subjectivist's guide to objective chance, in *Studies in Inductive Logic and Probability*, Vol. II, R. C. Jeffrey (ed.) (Los Angeles, CA: University of California Press, 1980), pp. 263–293. (Reprinted in [23], pp. 83–113.)

[23] D. Lewis, *Philosophical Papers*, Vol. II (Oxford: Oxford University Press, 1986).

[24] G. Ludwig, *Foundations of Quantum Mechanics I* (Berlin: Springer, 1983).
[25] D. N. Mermin, Simple unified form for the major no-hidden-variables theorems, *Phys. Rev. Lett.* **65**, 3373–3376 (1990).
[26] D. N. Mermin, Hidden variables and the two theorems of John Bell, *Rev. Mod. Phys.* **65**. 803–815 (1993).
[27] E. Schrödinger, Probability relations between separated systems, *Proc. Cambridge Philos. Soc.* **32**, 446–452 (1936).
[28] A. Shimony, Reality, causality, and closing the circle, in *Search for a Naturalistic World View, Volume I: Scientific Method and Epistemology* (Cambridge: Cambridge University Press, 2003), pp. 21–61.
[29] R. W. Spekkens, In defense of the epistemic view of quantum states: a toy theory, quant-ph/0401052 (2001).
[30] R. W. Spekkens, Evidence for the epistemic view of quantum states: a toy theory, *Phys. Rev. A* **75**, 032110 (2007).

10

Information, immaterialism, instrumentalism: Old and new in quantum information

Christopher G. Timpson

We live, we are told, in an information age. We are told this, perhaps, less often than once we were; but no doubt only because the phrase has become worn from use. If ours is an age of information, then quantum information theory is a field propitiously in tune with the spirit of the times: a rich and sophisticated physical theory that seeks to tame quantum mysteries (no less!) and turn them to ingenious computational and communication ends. It is a theory that hints, moreover, at the possibility of finally rendering the quantum *unmysterious*; or at least this is a conclusion that many have been tempted to draw.

Yet, for all its timeliness, some of the most intriguing of the prospects that quantum information science presents are to be found intertwining with some surprisingly old and familiar philosophical themes. These themes are immaterialism and instrumentalism; and in this chapter we shall be exploring how these old ideas feature in the context of two of the most tantalizing new questions that have arisen with the advent of this field. Does quantum information theory finally help us to resolve the conceptual conundrums of quantum mechanics? And does the theory indicate a new way of thinking about the world – one in which the material as the fundamental subject matter of physical theory is seen to be replaced by the immaterial: information?

The moral I will suggest is that it is only once the influence of these old ideas is explicitly recognized for what it is and treated accordingly that one can begin to hope for genuine new insights stemming from quantum information theory. Shannon, in the 1950s, warned against uncritical appeal to the concept of information [18]. We do well to heed his warning now.

10.1 Two thoughts

Why do our two tantalizing questions arise? Why should one think that quantum information theory – a branch of quantum mechanics intersecting with computer

Philosophy of Quantum Information and Entanglement, ed. A. Bokulich and G. Jaeger. Published by Cambridge University Press.

science and communication theory – might have any particular philosophical consequences in the first place? There are several reasons,[1] but the two most central to our concerns might be introduced in the following way.

Let us call the thought that information might be the basic category from which all else flows *informational immaterialism*. In this view, the new task of physics, foreshadowed in the development of quantum information theory, will be to describe the various ways in which information can evolve and manifest itself. Why might one be led to such a view? The thought could be as straightforward as this: we now have a fundamental – that is to say, a quantum – theory of information; so perhaps the fundamental theory of the world could just be about information (immaterial) rather than about things (material).[2]

Wheeler, with his "It from Bit" proposal is the cheerleader for this sort of view (it = physical thing; bit = information):

> No element in the description of physics shows itself as closer to primordial than the elementary quantum phenomenon . . . in brief, the elementary act of observer participancy . . . It from Bit symbolizes the idea that every item of the physical world has at bottom – at very deep bottom, in most instances – an immaterial source and explanation; that which we call reality arises in the last analysis from the posing of yes–no questions that are the registering of equipment evoked responses; in short, that all things physical are information-theoretic in origin and this is a *participatory universe*. ([26], pp. 3–5)

Or compare Steane:

> It now appears that information may have a much deeper significance. Historically, much of fundamental physics has been concerned with discovering the fundamental particles of nature and the equations which describe their motions and interactions. It now appears that a different programme may be equally important: to discover the ways that nature allows . . . *information* to be expressed and manipulated, rather than particles to move. ([20], pp. 120–121)

Finally, Zeilinger:

> So, what is the message of the quantum? I suggest we look at the situation from a new angle. We have learned in the history of physics that it is important not to make distinctions that have no basis – such as the pre-Newtonian distinction between the laws on Earth and those that govern the motion of heavenly bodies. I suggest that in a similar way, the

[1] See [23] for an introduction to the theory that seeks to emphasize various issues of philosophical interest.

[2] I make no claim that this is a good argument. Far from it. But it does seem to represent at least one strand of thought – largely unarticulated, perhaps – operative amongst those pursuing these idealist lines. Notice that it equivocates between two senses of the term "information," namely "information" as a technical term introduced by an information theory, quantum or classical; and "information" as the everyday semantic/epistemic term. These are importantly distinct (see [22–25] for discussion). To be the grounds for an immaterialist metaphysics "information" will need to refer to the semantic/epistemic concept: pieces of information would be the correlate – in modern terminology – of good old sense data, but "information" as it features in information theory has no direct link to any semantic or epistemic concept.

distinction between reality and our knowledge of reality, between reality and information cannot be made. [30]

Mixed in here, in the quotations from Wheeler and Zeilinger, is another important element in the immaterialist drive: strands of Copenhagen thought on the meaning of quantum mechanics. This is something we shall be returning to as we proceed.

The second thought arises perhaps more intuitively. It is very natural to think that the advent of quantum information theory might shed light on the conceptual troubles of quantum mechanics. After all, the *central* problem in quantum mechanics, the problem on which all else turns, is the measurement problem. Yet what is a measurement? Zurek sets the question up nicely:[3]

Quantum measurements are usually analysed in abstract terms of wavefunctions and Hamiltonians. Only a very few discussions of the measurement problem in quantum theory make an explicit effort to consider the crucial issue – the transfer of information. Yet obtaining knowledge is the very reason for making a measurement. [31]

So a measurement is an attempt to gain knowledge (information). But now we have a *quantum* theory of information: enlightenment is sure to follow! Or so the thought.

So much by way of introduction. We will begin the main discussion by exploring in more detail some of the ways in which it has been argued that appeal to the concept of information will aid our understanding of the basic conundrums in quantum mechanics: specifically the problems of measurement and non-locality. Hartle [10] illustrates a common strategy: if the quantum state is understood to represent information rather than an objective feature of the world, our troubles seem to disappear. However, I will suggest that this strategy proves problematic. It would seem either tacitly to invoke hidden variables, or to slide into a form of instrumentalism. But instrumentalism is not in itself a particularly edifying interpretive option: if this is all that appeal to information would amount to, we would not have succeeded in articulating a position of any interest. A further problem for the strategy can be noted: the factivity of the term "information" implies that the objectivity it was the express aim of the approach to avoid is inevitably reintroduced. It follows that, if one is to make any progress by associating the quantum state with some cognitive state, it must be the state of belief that is chosen, not that of knowledge.

One might take a different tack. It is possible to avoid the unedifying descent into instrumentalism by focusing instead on the question of whether information-theoretic principles might play the role of providing a perspicuous axiomatic basis for quantum mechanics, as some authors have urged (e.g., [6, 9]). Here we shall focus on Zeilinger's proposed information-theoretic foundational principle

[3] Although beware again of the possibility of equivocation between different senses of the term "information."

for quantum mechanics [29]. His hope is to explain the appearence of intrinsic randomness and entanglement in the theory; and ultimately to answer Wheeler's (1990) question "Why the quantum?" in a way congenial to the Bohrian intuition that the structure of quantum theory is a consequence of limitations on what can be said about the world. On consideration, however, this approach proves wanting, both formally and conceptually: appeal to the Foundational Principle cannot achieve the results desired.

We will close by exploring in more detail the links between Zeilinger's program, informational immaterialism, and certain strands of Copenhagen thought, specifically the remark infamously attributed to Bohr: "There is no quantum world." A good way of understanding what is going on in this remark is by viewing it as an example of semantic ascent. Thus understood, it becomes very clear that moves towards immaterialism are not supported by any such ascent.

10.2 The quantum state as information

Our two themes are immaterialism and instrumentalism. For better or worse, these are themes that have always been associated, more or less strongly, with the Copenhagen school of thought deriving from Bohr.[4] The notion of information has, in addition, often been appealed to by those working in this tradition. For this reason, Copenhagen-flavored interpretations have enjoyed something of a renaissance in recent years, on the back of quantum information theory (thereby bucking, to some extent, a contrary trend dating from the late 1960s toward broadly realist philosophies of science), since a quantum theory of information would seem to make such appeals to information more precise, and more scientifically respectable.

The thought typically proceeds by suggesting that, far from the central theoretical element of quantum theory – the quantum state – representing how things are in an external, objective world, it merely represents what information one has. References [12, 13, 26, 29] have all endorsed this kind of view. Hartle provides an excellent summary:

> The state is not an objective property of an individual system but is that information, obtained from a knowledge of how a system was prepared, which can be used for making predictions about future measurements.

[4] Needless to say, perhaps, it is very hard to discern any one thing that one might call *the* Copenhagen Interpretation of quantum mechanics. It is better to see a pattern of views centered around – and diverging in different ways from – Bohr's own. (For recent studies one might consult [3, 11].) Furthermore, of course, the exact nature of Bohr's own views (neo-Kantian? idealist? entity realist? patch-work realist?) is a matter of controversy. What is undeniable is that he took an instrumentalist view of the quantum formalism itself and had no straightforwardly realist metaphysics.

> . . . A quantum mechanical state being a summary of the observer's information about an individual physical system changes both by dynamical laws, and whenever the observer acquires new information about the system through the process of measurement. The existence of two laws for the evolution of the state vector . . . becomes problematical only if it is believed that the state vector is an objective property of the system . . . The "reduction of the wavepacket" does take place in the consciousness of the observer, not because of any unique physical process which takes place there, but only because the state is a construct of the observer and not an objective property of the physical system. ([10], p. 709)

Adopting this approach would seem to transform familiar, difficult, problems into non-problems. We may illustrate with two examples.

Consider one of the common formulations of the measurement problem: Wigner's friend [28]. Here we imagine two scientists, Wigner and friend, one of whom – the friend – is going to perform some quantum experiment and observe its outcome, whilst the other – Wigner – waits outside the (sealed) laboratory, unable to observe proceedings. How should we describe this scenario quantum mechanically? Suppose the experiment is a measurement of the z-component of spin of an electron prepared spin-up in the x-direction. Then the electron's initial state is the superposition

$$|\psi\rangle_{\text{initial}} = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle), \tag{10.1}$$

corresponding to no definite value of spin in the z-direction. On performing the measurement, Wigner's friend will see a definite outcome corresponding either to spin-up or to spin-down in the z-direction; accordingly he will assign one of the two spin states $|\uparrow\rangle$ or $|\downarrow\rangle$ to the post-measurement particle, corresponding to a definite z-spin value:

$$|\psi\rangle_{\text{final}} = |\uparrow\rangle, \text{ or } |\psi\rangle_{\text{final}} = |\downarrow\rangle. \tag{10.2}$$

But how will Wigner describe things? Since the electron-plus-apparatus-plus-friend-plus-lab consitutes a closed physical system, Wigner will describe the evolution of this system unitarily. Thus, according to him, the post-measurement state is not one in which the z-spin of the particle is definite, but is one in which the contents of the lab – electron, apparatus, and friend all included – is in one big superposition: the initial superposition of the spin is just amplified up to infect everything else:

$$\begin{aligned}|\Psi\rangle_{\text{lab,final}} = \frac{1}{\sqrt{2}}\big(&|\uparrow\rangle_{\text{electron}}|\text{“up”}\rangle_{\text{apparatus}}|\text{sees “up”}\rangle_{\text{friend}} \\ &+ |\downarrow\rangle_{\text{electron}}|\text{“down”}\rangle_{\text{apparatus}}|\text{sees “down”}\rangle_{\text{friend}}\big).\end{aligned} \tag{10.3}$$

Who is right? Does Wigner's friend see a definite outcome, or is he left suspended in limbo until Wigner opens the door to say hello? Do we need to appeal to collapse

to reconcile the two views? If so, how, when, and why does collapse occur? Why *isn't* Wigner right to treat a closed physical system as unitarily evolving? And so on. These are the familiar kinds of worries.[5]

The informational approach aims to undercut this dialectic neatly. The thought is that there need be no disagreement between Wigner and friend. If the quantum state does not represent how things are in the world but what information somebody possesses, then Wigner can assign a state like (10.3) and his friend can assign one of the states (10.2) without them thereby making contradictory or contrary statements. They are not disagreeing about how things are, they merely have access to different information – Wigner outside the laboratory and his friend inside. When Wigner gains more information by intrepidly entering the lab, he will update his state accordingly, but, since that update does not correspond to a change in anything in the world, it is not a mysterious change, or one in need of explanation. Thus the argument proceeds.

The second illustrative example is non-locality. Take a familiar EPR scenario. Two parties, Alice and Bob, are space-like separated and each possess one half of an entangled pair of particles, e.g., one of a pair of spin-$\frac{1}{2}$ systems in the singlet state:

$$|\psi\rangle_{\mathrm{AB}} = \frac{1}{\sqrt{2}}(|\uparrow\rangle_{\mathrm{A}}|\downarrow\rangle_{\mathrm{B}} - |\downarrow\rangle_{\mathrm{A}}|\uparrow\rangle_{\mathrm{B}}). \tag{10.4}$$

Alice will then perform a measurement on her system. Prior to measurement both Alice and Bob will assign the maximally mixed state to both of the subsystems of the pair:

$$\rho_{\mathrm{A}} = \rho_{\mathrm{B}} = \frac{1}{2}|\uparrow\rangle\langle\uparrow| + \frac{1}{2}|\downarrow\rangle\langle\downarrow| = \frac{1}{2}\mathbf{1}. \tag{10.5}$$

The standard story [7] is that the effect of Alice's measurement will be to change the state of Bob's system *instantaneously* and *at a distance*. If she measured in the z-spin basis, for example, obtaining the "up" outcome, the state ascribed to Bob's system will now be $|\downarrow\rangle_{\mathrm{B}}$ rather than $(1/2)\mathbf{1}$.

Again, the informational approach will suggest that this conclusion of action at a distance is predicated on a false assumption: that the quantum states represent how things are in the world. If they do not, then we no longer have action at a distance. Post-measurement, Alice will assign a different state to Bob's system than he himself does, for example the pure state $|\downarrow\rangle_{\mathrm{B}}$, while he still assigns the mixed state $(1/2)\mathbf{1}$; but these differences merely represent differences in the

[5] Recall that Wigner's own view was that one *does* need to invoke a process of collapse, but it is in some sense a partially non-physical process, being brought about by the effect of conscious mind (and not just one's own mind, as Wigner preferred not to conceive solipsistically of his friend hanging in limbo if left to his own devices).

information the two possess. The change in the global state from the singlet to the post-measurement state $|\uparrow\rangle_A|\downarrow\rangle_B$ (or equivalently the change in the state of Bob's system from $(1/2)\mathbf{1}$ to $|\downarrow\rangle_B$) does not, in this view, represent a change in the world, a change in Bob's system, but merely an update of Alice's information; *a fortiori* the change does not involve non-locality. Bob will have no opportunity to update his information about his system until he later meets up with Alice and she reports the outcome of her measurement; but the fact that they assign different states need not mean that they are disagreeing about how the world is.

10.3 Against "information"

So we can see why it might seem appealing to call on the notion of information when trying to make sense of the quantum state. Some thorny problems seem to be dissolved, revealed to be the result of a jejune literalism about the quantum formalism. But things are not really so straightforward. With characteristic perspicacity and concision, John Bell put his finger right on the nub of the central problem with this approach.

"Information" features as one of the bad words on Bell's famous list of terms having "no place in a formulation with any pretence to physical precision" [2]. Why? Bell indicated the source of his disquiet by posing two questions: if information, then information about *what*? And *whose* information?

I take the first of these to be the most pressing. It presents the informational approach with a troublesome dilemma. If the quantum state represents one's information, there seem to be only two sorts of answer possible to "information about what?":

(1) information about what the outcome of experiments will be; or
(2) information about how things are with a system prior to measurement.

Neither of these, I suggest, can happily be adopted by our would-be informationist. Consider answer (2). The information concerns properties of a system that are possessed prior to measurement and aren't described by the quantum state (in this case because the state doesn't have a world-describing role). What is the more familiar name for such properties? Hidden variables, of course. But recall: the whole point of taking the quantum state as information was to mollify its bad behavior, its jumping here and there we know not when, its non-local collapse. But if to do that we need to introduce hidden variables – to be what it is that the state represents information about – then we are even worse off than we were before. Because, as we all know, hidden variables have to be *very badly behaved indeed* in quantum

mechanics (non-locality, contextuality). Thus it would seem to be self-defeating for the informationist to take option (2).[6]

Let's turn then to the first answer. If the information the state represents is information about what the results of experiments will be, then the difficulty now is to say anything interesting that doesn't simply slide into instrumentalism. Instrumentalism, of course, is the general view that scientific theories do not seek to describe the laws governing unobservable things, but merely function as devices for predicting the outcomes of experiments. An instrumentalist view of the quantum state understands the state merely as a device for calculating the statistics of measurement outcomes. How is the current view any different, apart from having co-opted the vogue term "information"? If all that appeal to information were ultimately to amount to is a form of instrumentalism then we would not have achieved a particularly interesting – and certainly not a novel – interpretive doctrine. It would be an error to let a superficial repackaging in fancy wrappings convince one that a product was worth buying after all. To the founding fathers of quantum mechanics, instrumentalism might conceivably have seemed a progressive epistemological doctrine, but that can scarcely be said to be the case now. As an option for interpreting quantum mechanics it arguably amounts more to a refusal to ask questions than to taking quantum mechanics seriously.

10.3.1 The problem of factivity

But perhaps we have been a little too precipitate in our analysis. Might there be a subtlety of the informational approach that we have so far missed? Possibly. It is perhaps useful to highlight a subdivision in instrumentalist views that I have been glossing over.

Consider once more what might be called standard instrumentalism about the quantum state. This works as an interpretation of quantum mechanics (insofar as it works, that is) by witholding any descriptive claims at the level of individual systems. It restricts itself to making claims only about measurement results on ensembles of systems. (So in this view it would be a badly posed question to ask in quantum mechanics something like how does an individual electron travel in a two-slit experiment? One can only ask about what observable *results* one might expect to see for *very many* electrons.)

[6] A caveat. If one adopted an informational view of the state not in order to address the measurement problem; and not in order to relieve problems over non-locality; if one could argue that it was *natural* for *quite other reasons*, perhaps, to take the state to represent information, then one might not be so moved by this objection; and one might willingly embrace the charge that one was dealing with hidden variables. See [19]. Of course, one must then admit that it's not really the notion of information that is doing any of the interesting work.

So what if we were to insist that, in contrast with this, the distinctive job of information talk is to allow one to talk about *individual* systems, not just ensembles (look again at Hartle's wording above). Perhaps *this* is what would make the informational approach a worthwhile novel approach. Sadly, this approach faces a decisive objection.

Let us begin by noting that descriptions of the quantum state in terms of a person's knowledge or information will typically involve what might be called *mixed* ascriptions. That is, they will involve both the everyday semantic/epistemic concept of information and, at the same time, the distinct technical concept of information introduced in information theory. We see this when we recognize that one will need to answer the question about *what* information the state represents (Bell's question again); and will answer by talking of information *that* p or *about* q, both locutions signaling the everyday concept. At the same time, one might be interested in *how much* information the state represents, a phrase typically signaling the technical concept.[7]

However, once we have the everyday concept of information in play, we need to recognize that the term "information" is, just as the term "knowledge" is, *factive*. That is, having the information that p entails that p is the case. Just as I can't know that p unless it is true that p, no more can I have the information that p unless p. The difficulty that this presents for those wishing to understand the quantum state of an individual system as information is that this factivity entails just the sort of objectivity it was the original aim of the approach to avoid.

The standard instrumentalist does not face quite this problem. The problems associated with measurement and non-locality are avoided by remaining at the level of statistics only. One doesn't describe individual systems at all and collapse will not be thought of as a real process, merely a change in what statistics one will expect. But, for the proponent of the quantum state as information about individual systems, the essence of *their* approach, as we have seen above, is that different agents can assign different states to a given system because they have different information regarding it, but without disagreeing. In the Wigner's-friend scenario, there would be no mysterious collapse, both agents simply ascribe a different state to the system being measured; and there is no one correct state that is an objective property of the system. Similarly with the EPR case.

But the factivity of information and knowledge put paid to these forms of argument: if the quantum state represents what one knows, or what information one has, then things have to be as they are known to be. If I know what the probability

[7] As noted above, consult [22–25] for more on the distinction between these concepts. It is plausible to suppose that at least part of the common failure to distinguish sufficiently between the everyday and technical concepts of information may be traced to the existence of mixed contexts in which the two distinct concepts are employed in the same breath; and, confusingly, using the same word.

distributions for measurements on a system are, then they must objectively be thus-and-so. It is a matter of right or wrong determined by what the properties of the system are. Or again, if Alice performs a measurement on her half of the entangled pair and therefore subsequently knows the pure state of Bob's system, then his system objectively has to be in that state. It is now a determinate matter of fact that some measurement at Bob's location will have a definite outcome whereas before it didn't. We are forced once more, like it or not, to talk about objective properties that systems possess; moreover, objective properties that can be changed at a distance.

So it seems that this approach runs aground. Adopting "information" does not, after all, free us from the objectivity that was causing the trouble in the first place. If there is to be any mileage, therefore, in approaches that analyze the quantum state in terms of cognitive states, one can't choose knowledge; and one must drop information. The state to choose would instead be *belief*, for believing that p does not entail p. For an approach that does just this, see the quantum Bayesianism of Caves, Fuchs, and Shack (e.g., [5, 8, 24]).

10.4 If not instrumentalism, axiomatics instead?

Let us change tack. Steering between the horns of the dilemma of hidden variables versus instrumentalism proves difficult, maybe impossible. Perhaps we would do better to explore instead the possibility that the concept of information might have a role to play in rendering quantum mechanics more perspicuous via a suitable axiomatization in information-theoretic terms. We shall focus on Zeilinger's version of this project because it links in interesting ways with our two overall themes.

Zeilinger [29] suggests that we can render quantum mechanics more readily intelligible when we see how various of its fundamental and distinctive features can be derived from a simple principle:

Foundational Principle: *an elementary system represents the truth value of one proposition.*

This is also expressed as the claim that elementary systems carry only one bit of information. The idea is that we have here something akin to the principle of relativity in special relativity, or to the principle of equivalence in general relativity; namely a simple and intuitively compelling principle that plays a key role in deriving the structure of the theory. In particular, Zeilinger argues that the principle allows us to understand both where the *irreducible* randomness of quantum mechanics and where entanglement come from. (Both, surely, centrally quantum – and centrally mysterious – features.) Elementary systems are components that are arrived at as the end point of a process of analysis. One begins by analyzing a

composite system into smaller component parts. Zeilinger then suggests that it is natural to assume that each consituent system will require fewer propositions for its description than the composite[8] and we might furthermore expect to reach a limit:

> . . . the limit is reached when an individual system represents the truth value to one proposition only. Such a system we call an *elementary system*. ([29], p. 635)

We might term the propositions in question *elementary propositions* too. Notice that the Foundational Principle is now revealed to be a tautology, or perhaps an analytic truth: a definition of "elementary system." One might be surprised that much could then follow from this on its own. This attitude proves warranted.

Zeilinger means something quite specific by "proposition." As in the quantum-logical tradition, it means something that represents an experimental question; and a truth-value assignment to a proposition corresponds to a yes/no answer to the experimental question. This allows us to formulate the principle more perspicuously as the following claim:

The state of an elementary system specifies the answer to a single yes/no experimental question.

One might not agree that this is a suitably unrestricted conception of the state of a system (compare this with the de Broglie–Bohm theory, for example, whose elements of holism and contextuality render this conception quite inapposite) but let us put this worry to one side and focus instead on what work the Foundational Principle is supposed to do within its own domain. The proposed explanation of the genesis of quantum randomness is very simple. Given the Foundational Principle,

> . . . an elementary system cannot carry enough information to provide definite answers to all questions that could be asked experimentally. ([29], p. 636)

Those questions which don't receive a definite answer must then receive a random answer; and, furthermore, that randomness *must* be irreducible, since, if it *could* be reduced to hidden properties, then the system would really be carrying more than one bit of information, in violation of the principle (assuming that the system was in fact elementary).

For the explanation of entanglement, suppose we have N systems and they have N bits of information associated with them. Entanglement results when all those

[8] Is it? Only relative to a fixed system of concepts adequate to describe all levels of physical complexity; i.e., in which one begins with elementary propositions describing basic objects; and more complex objects are described by truth-functional combinations of these elementary propositions. (Consider: one could plausibly maintain that it takes fewer propositions to describe a *table* adequately than it does to describe an electron. Doesn't the sheer effort involved in science show that it typically gets *harder* to describe things the smaller they are?) Zeilinger's approach here bears marked similarities to Wittgenstein's views in the *Tractatus Logico-Philosophicus*.

Table 10.1 *The four Bell states, a basis of maximally entangled two-party states*

$\|\phi^+\rangle = (1/\sqrt{2})(\|\uparrow\rangle\|\uparrow\rangle + \|\downarrow\rangle\|\downarrow\rangle)$
$\|\phi^-\rangle = (1/\sqrt{2})(\|\uparrow\rangle\|\uparrow\rangle - \|\downarrow\rangle\|\downarrow\rangle)$
$\|\psi^+\rangle = (1/\sqrt{2})(\|\uparrow\rangle\|\downarrow\rangle + \|\downarrow\rangle\|\uparrow\rangle)$
$\|\psi^-\rangle = (1/\sqrt{2})(\|\uparrow\rangle\|\downarrow\rangle - \|\downarrow\rangle\|\uparrow\rangle)$

bits of information are used up in specifying *joint* properties of the system, rather than individual properties, or, more generally, when more information is in the joint properties than is possible classically [4].

To illustrate the claim about entanglement we may use the case of two qubits. Consider the four maximally entangled bipartite quantum states known as the *Bell states* (Table 10.1). Each of these is a joint eigenstate of the observables $\sigma_x \otimes \sigma_x$ and $\sigma_y \otimes \sigma_y$. From the Foundational Principle, only two bits of information are associated with our two systems, i.e., the states of these systems can specify the answer to two experimental questions only. If the two questions whose answers are specified are "Are both spins in the same direction along x?" $((1/2)(\mathbf{1}\otimes\mathbf{1}+\sigma_x\otimes\sigma_x))$ and "Are both spins in the same direction along y?" $((1/2)(\mathbf{1}\otimes\mathbf{1}+\sigma_y\otimes\sigma_y))$, then we end up with a maximally entangled state. The answer "yes, yes" would give use the state $|\phi^+\rangle$; "yes, no", the state $|\psi^+\rangle$; "no, yes", the state $|\phi^-\rangle$; and "no, no", the singlet state $|\psi^-\rangle$.

By contrast, if the two questions had been "Are both spins in the same direction along x?" and "Is the spin of particle 1 up along x?," the information would not have all been used up specifying *joint* properties and we would have instead a product state, a joint eigenstate of $\sigma_x \otimes \sigma_x$ and $\sigma_x \otimes \mathbf{1}$. "Yes, yes" would give us the state $|\uparrow\rangle|\uparrow\rangle$; "yes, no" would give us $|\downarrow\rangle|\downarrow\rangle$; "no, yes" the state $|\uparrow\rangle|\downarrow\rangle$; and "no, no" the state $|\downarrow\rangle|\uparrow\rangle$. The way the information is distributed over the systems is crucial in determining whether we have an entangled or a non-entangled state.

At first glance, these explanations might appear to have something going for them, but only at first glance. Unfortunately they suffer from a deep flaw. No attention has been paid to the structure of the set of experimental questions on individual and joint systems, yet it is precisely this which is essential to the appearance of randomness and entanglement. The Foundational Principle places no constraints on the set of experimental questions at all, so it cannot do the job of explaining the existence of quantum randomness and entanglement.[9]

[9] These objections were presented in [21].

Consider: irreducible randomness would arise only if there were more experimental questions to be asked of an elementary system than its most detailed state description would provide an answer for.[10] But what determines how many experimental questions there are and how they relate to one another? Certainly not the Foundational Principle. The principle doesn't explain why it is that, having given the finest-grained description of the system that is possible, any space for randomness still remains. Why isn't the one bit enough on its own, for example? In the quantum case, because the set of experimental questions is in one-to-one correlation with projectors onto a complex Hilbert space and the simplest non-trivial state space is two-dimensional. But that is the structure we are supposed to be deriving, it is not something we can help ourselves to. Why should the set of experimental questions be structured like *that*? Compare this with the state space of a classical Ising-model spin: these objects have only two states: up or down; here the one bit we are allowed per system is quite sufficient to answer all experimental questions that could be asked. The Foundational Principle is clearly powerless to distinguish between the two cases.

The case of entanglement is similar. If we return to the starting point and consider our N elementary systems, all that the Foundational Principle tells us regarding these systems is that their individual states specify the answer to a single yes/no question concerning each system individually. There is, as yet, no suggestion of how this relates to joint properties of the combined system. Some assumption needs to be made before we can go further. For instance, we need to enquire whether there are supposed to be experimental questions regarding the joint system that can be posed and answered that are not equivalent to questions and answers for the systems taken individually. (We know that this will be the case, given the structure of quantum mechanics, but again we are not allowed to *assume* this structure, if we are engaged in a foundational project.[11]) If this *is* the case then there can be a difference between the information associated with correlations (i.e., regarding answers to questions about joint properties) and the information regarding individual properties. But then we need to ask the following question: why is it that there exist sets of experimental questions to which the assignment of truth values is not equivalent to an assignment of truth values to experimental questions regarding individual systems?

Because such sets of questions exist, more information can be "in the correlations" than in individual properties. Stating that there is more information in correlations than in individual properties is then to report that such sets of non-equivalent

[10] Suppose, for simplicity, that we have some kind of independent access to the notion of an elementary system, so we can tell whether a system really is supposed to be elementary or not.

[11] To illustrate, a simultaneous truth-value assignment for the experiments $\sigma_x \otimes \sigma_x$ and $\sigma_y \otimes \sigma_y$ cannot be reduced to one for experiments of the form $\mathbf{1} \otimes \mathbf{a} \cdot \boldsymbol{\sigma}$, $\mathbf{b} \cdot \boldsymbol{\sigma} \otimes \mathbf{1}$.

questions exist, *but it does not explain why they do so*. However, it is surely this that demands explanation – why is it not simply the case that all truth-value assignments to experimental questions are reducible to truth-value assignments to experimental questions regarding individual properties, as they are in the classical case? That is, why does entanglement exist? In the absence of an answer to the question when posed in this manner, the suggested explanation following from the Foundational Principle seems dangerously close to the vacuous claim that entanglement results when the quantum state of the joint system is not a separable state.

As it stands, the Foundational Principle is wholly unsuccessful. Might we be able to salvage something from the approach, however? Perhaps if we were to add further axioms that entailed something about the structure of the set of experimental questions, progress could be made. A possible addition might be a postulate Rovelli adopts: *it is always possible to acquire new information about a system* [17]. One wouldn't be terribly impressed by an explanation of irreducible randomness invoking the Foundational Principle and this postulate, however, since it would look rather too much like putting the answer in by hand. But there might be other virtues of the system to be explored.

This consideration raises a final point. Evidently more axioms need to be added if we are to derive any useful work from the Foundational Principle. But recall that to provide a perspicuous information-theoretic axiomatization, we will need intuitively compelling axioms (or at least *some* of them) that also play a substantial role. It would not do simply to add any old axioms that would have the effect of recovering the correct structure of experimental questions, otherwise there is no explanatory gain to be had. Recall that there has been considerable progress in the quantum-logical tradition of providing axiomatizations of quantum mechanics (see, e.g., [1] for a succinct review), but it is not clear that *these* approaches render quantum mechanics any less mysterious or any more intuitively understandable. It is not clear, in any case, that this is really their purpose.

10.5 Why the quantum?

We have seen that there are formal difficulties with Zeilinger's approach; let us now consider some of its more philosophical underpinnings. There are clear affinities between the views Zeilinger expresses and Wheeler's It from Bit proposal. For example, Zeilinger states that, as he intends it, the statement

> that a system "represents" the truth value of a proposition . . . only implies what can be said about possible measurement results. ([29], p. 635)

Moreover, in his view, the results of measurement do not pertain to an externally existing mind-independent world; rather his view is an immaterialist one:

properties are assigned to objects only on the basis of observation and are held only as long as they do not contradict further observation; and "In fact, the object therefore is a useful construct connecting observations" ([29], p. 633).[12] Extreme subjectivism is kept in check by the requirement that there be intersubjective agreement between different agents' "mentally constructed objects" ([29], p. 634). Clearly, this kind of immaterialist setting would make the Foundational Principle more plausibly appear a good starting point for theorizing.

So Zeilinger and Wheeler seem to share an immaterialist metaphysics. Of particular interest, however, is a striking passage in which Zeilinger suggests that his Foundational Principle might provide an answer to Wheeler's question "Why the quantum?" in a way that chimes with the Bohrian thought that the structure of quantum theory is a consequence of limitations on what can be said about the world:

> The most fundamental viewpoint here is that the quantum is a consequence of what can be said about the world. Since what can be said has to be expressed in propositions and since the most elementary statement is a single proposition, quantization follows if the most elementary system represents just a single proposition. ([29], p. 642)

Of course, we have already, in effect, seen that there is a crucial non-sequitur here. Quantization follows only if the propositions are projection operators on a complex Hilbert space. Why the world has to be described in *that* way is the question that would really need to be answered in answering Wheeler's question; and the Foundational Principle does not help us. But it is interesting to delve a little further into why this non-sequitur is present. Reflect on the similarity between Zeilinger's statement and that famously attributed to Bohr by Petersen:

> There is no quantum world. There is only an abstract quantum physical description. It is wrong to think that the task of physics is to find out how nature *is*. Physics concerns what we can say about nature. ([14], p. 12)

The last sentence is particularly pertinent: "Physics concerns what we can say about nature." Compare this, again, with another statement of Zeilinger's, ". . . what can be said about Nature has a constitutive contribution on what can be 'real.' " (reported in [9], p. 615).

I think we find in these sentiments a crucial strand contributing to the thought that the rise of quantum information theory supports an informational immaterialism. If quantum mechanics reveals that the true subject matter of physics is what can be said, rather than how things are, then this seems very close to saying that

[12] It is perhaps unnecessary to add that, if the foregoing is supposed to be an argument for the immaterialist position, it is an extremely weak one, failing as it does, for example, to distinguish between the grounds on which one might assert a proposition and what would thereby have been asserted.

what is fundamental is the play of information across our psyches. The development of a quantum theory of information merely exacerbates this thought stemming from the Copenhagen tradition.

However, it is important to recognize that there is a very obvious difficulty with the thought that what can be said provides a constitutive contribution to what can be real and that physics correspondingly concerns what we can say about nature. Simply reflect that some explanation needs to be given of where the relevant constraints on what can be said come from. Surely there could be no other source for these constraints than the way the world actually is – it can't *merely* be a matter of language.[13] It is because of the unbending nature of the world that we find the need to move, for example, from classical to quantum physics; that we find the need to revise our theories in the face of recalcitrant experience. Zeilinger and Bohr (in the quotation above) would thus seem to be putting the cart before the horse, to at least some degree. Schematically, it's the way the world is (independently of our attempted description or systematization of it) that determines what can usefully be said about it, and that ultimately determines what sets of concepts will prove most appropriate in our scientific theorizing. It is failure to recognize this simple truth that accounts, I suggest, for the otherwise glaring non-sequitur in the proposed answer to "Why the quantum?." One can't expect a substantive empirical truth (e.g., about the correct structure of the set of experimental questions) to follow from a simple definitional statement like the Foundational Principle.

Another point can be drawn from the Petersen quotation. With its focus on the level of physical *description* and what can be *said* about nature (as opposed to how nature is) this passage can be seen to provide us with an example of what is often known as *semantic ascent*.

Semantic ascent is the move from what Carnap called the material mode to the formal mode, that is, roughly speaking, from talking about things to talking about words. As Quine says, "*semantic ascent* ... is the shift from talking in certain terms to talking about them" ([16], p. 271). Bohr, it would seem, would have us ascend from the level of using words within our theory to the level of describing our descriptions. This, the suggestion is, is the true task of physics.

[13] Of course, what statements can be made depends on what concepts we possess; and, trivially, in order to succeed in making a statement, one needs to obey the appropriate linguistic rules. But the point at issue is what can make one set of concepts more fit for our scientific theorizing than another? For example, why do we have to replace commuting classical physical quantities with non-commuting quantum observables? As Quine perspicuously notes "... truth in general depends on both language and extra-linguistic fact. The statement 'Brutus killed Caesar' would be false if the world had been different in certain ways, but it would also be false if the word 'killed' happened to have the sense of 'begat.' "([15], p. 36). The world is required to provide the extra-linguistic component that will make one set of concepts more useful than another; furthermore, without an extra-lingustic component to truth, we could only ever have analytic truths – and that would no longer be physics.

But does such semantic ascent really achieve very much here? Far from it. It is true, but entirely trivial, that our subject matter will not be the world if we semantically ascend. It would be just as true in classical mechanics, say, as in quantum mechanics. There is indeed a sense in which there would be no quantum world at the level of our interest, but only because we are talking about words rather than the world; talking *about* various terms rather than *in* them.

Crucially, the fact that one has ascended doesn't make the level one has ascended from go away: notice that one can always force descent by asking "So, what was said?." It might perhaps be felt that here lies the real import of the Bohr quote and what serves to distinguish the quantum from the classical case. In the quantum case, we might be supposed to imagine that one *can* intelligibly kick away the lower level, having made the semantic ascent. But such a suggestion ("vertiginous semantic ascent," as it might be called) is in fact incoherent. It would amount to the claim that the "descent" question "So, what was said?" becomes unintelligible, but this would entail that the terms under discussion have to become entirely devoid of meaning, and as such they would have no role whatsoever in physics.

The upshot is that we can't shirk any of the problems of interpreting quantum mechanics by indulging in semantic ascent. It doesn't remove us from the fray or amount to an interpretation of quantum mechanics in itself. The world doesn't disappear because we may be talking about the terms in which we describe it. The interpretational questions that have always plagued quantum theory concern what stance should be taken to claims made *using* the terms within a theory; and all the usual options (realism, instrumentalism, and hybrids thereof) will remain open irrespective of ascent.

If this reading I have suggested is indeed the most intelligible reading of the Bohr quotation then it becomes clear that "there is no quantum world" and "physics concerns what we can say about nature" are not after all immaterialist mantras. Rather they are truistic consequences of an innocuous semantic ascent. In fact it's hard to see how they could be anything else while retaining the least hint of plausibility.

10.6 Conclusion

My aim in this essay has been to clear the ground a little. Quantum information theory is indeed a rich and intriguing subject for philosophical study, but, if we are to be sensitive to what new consequences it may have for our understanding of quantum mechanics in particular and physics in general, then we do better if we are able to separate new from old; and if we turn a suitably skeptical eye toward the claims of our familiar pair instrumentalism and immaterialism.

Elsewhere I have distinguished direct from indirect approaches to securing a philosophical dividend from quantum information theory. Among the direct approaches we can count taking the quantum state to be information and taking quantum information to support informational immaterialism. More interesting and plausible than these proposals, I suggest, are the indirect approaches. Amongst these are attempts such as Zeilinger's to learn something useful about the structure or axiomatics of quantum theory by reflecting on quantum information-theoretic phenomena; approaches that might look to quantum information theory to provide new analytic tools for investigating that structure; and approaches that look to suggested constraints on the power of computers, including quantum computers, as potential constraints on new physical laws. Whilst Zeilinger's particular program suffers from the rather severe problems we have seen and is tangled up with instrumentalist and immaterialist threads of broadly Copenhagen origin, in general, the indirect approaches look by far the most promising potential sources of new insights stemming from quantum information theory. By contrast the two direct approaches that we have been considering are not in good shape. The informational approach to the quantum state seems unable to survive the hidden-variables/instrumentalism dilemma; and the thought that quantum information theory does lend support to a form of immaterialism really seems to have very little to commend it.

Acknowledgments

This work, which draws on material discussed in [22, 25], was supported by a research leave award from the UK Arts and Humanities Research Council, for the project Information in Physics. I would also like to thank the Department of Philosophy at the University of Leeds for providing me with a period of matching leave.

References

[1] D. Aerts and S. Aerts, Towards a general operational and realistic framework for quantum mechanics and relativity theory, in *Quo Vadis Quantum Mechanics*, Elitzur, A., Dolev, S., and Kolenda, N. (eds.), (Berlin: Springer, 2005), pp. 152–207.

[2] J. S. Bell, Against "measurement," *Phys. World*. August, pp. 33–40 (1990). Reprinted in J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics*, 2nd Edn. (Cambridge: Cambridge University Press, 2004), pp. 213–231.

[3] M. Beller, *Quantum Dialogue: The Making of a Revolution* (Chicago, IL: Chicago University Press, 1999).

[4] C. Brukner, M. Zukowski, and A. Zeilinger, The essence of entanglement, arXiv:quant-ph/0106119 (2001).

[5] C. M. Caves, C. A. Fuchs, and R. Schack, Subjective probability and quantum certainty, *Stud. Hist. Philos. Mod. Phys.* **38**, 255–274 (2007).
[6] R. Clifton, J. Bub, and H. Halvorson, Characterizing quantum theory in terms of information theoretic constraints, *Foundations Phys.* **33**, 1561 (2003); arXiv:quant-ph/0211089 (2002).
[7] A. Einstein, B. Podolsky, and N. Rosen, Can quantum mechanical description of physical reality be considered complete?, *Phys. Rev.* **47**, 777–780 (1935).
[8] C. A. Fuchs, Quantum mechanics as quantum information (and only a little more), in *Quantum Theory: Reconsideration of Foundations*, A. Khrennikov (ed.) (Växjö: Växjö University Press, 2002); arXiv:quant-ph/0205039 (2002).
[9] C. A. Fuchs, *Notes on a Paulian Idea: Foundational, Historical, Anecdotal and Forward Looking Thoughts on the Quantum (Selected Correspondence)* (Växjö: Växjö University Press, 2003); arXiv: quant-ph/0105039 (2001).
[10] J. B. Hartle, Quantum mechanics of individual systems, *Am. J. Phys.* **36**, 704–712 (1968).
[11] D. Howard, Who invented the "Copenhagen Interpretation"? A study in mythology, *Philos. Sci.* **71**, 669–682 (2004).
[12] N. D. Mermin, Whose knowledge?, in R. Bertlmann and A. Zeilinger (eds.) *Quantum (Un)speakables: Essays in Commemoration of John S. Bell*, (Berlin: Springer, Berlin, 2001); arXiv:quant-ph/0107151 (2001).
[13] R. Peierls, In defence of "Measurement," *Phys. World* January, pp. 19–20 (1991).
[14] A. Petersen, The philosophy of Niels Bohr, *Bull. Atomic Scientists*, **19**(7), 8–14 (1963).
[15] W. V. O. Quine, *From a Logical Point of View*, 2nd edn. (Harvard, MA: Harvard University Press, 1953).
[16] W. V. O. Quine, *Word and Object*, (Cambridge, MA:MIT Press, 1960).
[17] C. Rovelli, Relational quantum mechanics, *Int. J. Theor. Phys.* **35**, 1637–1678 (1996).
[18] C. E. Shannon, The bandwagon, *IRE Trans. Information Theory*, **32**(1), (1956).
[19] R. Spekkens, Evidence for the epistemic view of quantum states: a toy theory, *Phys. Rev. A* **75**, 032110 (2007).
[20] A. Steane, Quantum computing, *Rep. Prog. Phys.* **61**, 117–173 (1998).
[21] C. G. Timpson, On a supposed conceptual inadequacy of the Shannon information in quantum mechanics, *Stud. Hist. Philos. Mod. Phys.* **34**, 441–468 (2003).
[22] C. G. Timpson, *Quantum Information Theory and the Foundations of Quantum Mechanics*, Ph.D. thesis, University of Oxford, arXiv:quant-ph/0412063 (2004).
[23] C. G. Timpson, Philosophical aspects of quantum information, *The Ashgate Companion to the New Philosophy of Physics*, Rickles (ed.) (Ashgate 2008), pp. 197–261.
[24] C. G. Timpson, Quantum Bayesianism: A study, *Stud. Hist. Philos. Mod. Phys* **39**, 579–609 (2008).
[25] C. G. Timpson, *Quantum Information Theory and the Foundations of Quantum Mechanics* (Oxford: Oxford University Press, forthcoming).
[26] J. A. Wheeler, Information, physics, quantum: the search for links, in *Complexity, Entropy and the Physics of Information*, W. Zurek, (ed.) (Redwood City, CA: Addison-Wesley, 1990), pp. 3–28.
[27] J. A. Wheeler and W. H. Zurek (eds.), *Quantum Theory and Measurement* (Princeton, MA: Princeton University Press, 1983).

[28] E. P. Wigner, Remarks on the mind–body question, In Good, I. J., editor, *The Scientist Speculates*, I. J. Good (ed.) (London: Heinemann, 1961), pp. 284–302. Reprinted in [27], pp. 168–181.
[29] A. Zeilinger, A foundational principle for quantum mechanics, *Foundations Phys.* **29**, 631–643 (1999).
[30] A. Zeilinger, The message of the quantum, *Nature* **438**, 743 (2005).
[31] W. Zurek (ed.), *Complexity, Entropy and the Physics of Information* (Redwood City, CA: Addison-Wesley, 1990).

Part IV

Quantum communication and computing

11

Quantum computation: Where does the speed-up come from?

Jeffrey Bub

11.1 Introduction

Discussions of quantum-computational algorithms in the literature refer to various features of quantum mechanics as the source of the exponential speed-up relative to classical algorithms: superposition and entanglement, the fact that the state space of n bits is a space of 2^n states while the state space of n qubits is a space of 2^n dimensions, the possibility of computing all values of a function in a single computational step by "quantum parallelism," or the possibility of an efficient implementation of the discrete quantum Fourier transform. Here I propose a different answer to the question posed in the title, in terms of the difference between classical logic and quantum logic, i.e., the difference between the Boolean classical event structure and the non-Boolean quantum event structure. In a nutshell, the ultimate source of the speed-up is the difference between a classical disjunction, which is true (or false) in virtue of the truth values of the disjuncts, and a quantum disjunction, which can be true (or false) even if none of the disjuncts is either true or false.

In the following, I will discuss the information-processing in Deutsch's XOR algorithm [6] (the first genuinely quantum algorithm) and related period-finding quantum algorithms (Simon's algorithm [14, 15] and Shor's factorization algorithm [12, 13]). It is well known that these algorithms can be formulated as solutions to a hidden-subgroup problem (see [9, 10]). Here the salient features of the information-processing are presented from the perspective of the way in which the algorithms exploit the non-Boolean logic represented by the projective geometry (the subspace structure) of Hilbert space.

A particular period partitions the domain of a periodic function – the input values for the algorithm – into mutually exclusive and collectively exhaustive subsets. Distinguishing the period from alternative possible periods amounts to distinguishing the corresponding partition from alternative possible partitions. A classical algorithm requires the evaluation of the function for a subset of input values

Philosophy of Quantum Information and Entanglement, ed. A. Bokulich and G. Jaeger. Published by Cambridge University Press.

to determine the partition – a computational task that involves an exponentially increasing number of steps as the size of the input increases. The trick in Simon's quantum algorithm, as we will see below, is to represent the alternative possible partitions by Hilbert-space subspaces that are orthogonal except for overlaps or intersections. Each subspace is spanned by states of the input register representing the different subsets in the associated partition. A measurement in the computational basis can provide sufficient information to identify the subspace containing the state (after a suitable transformation) and hence the partition associated with the period without evaluating the function at all (in the sense of producing a value in the range of the function for a value in its domain). The algorithm generally has to be run several times because the measurement might be inconclusive, corresponding to an outcome associated with the overlap region, but achieves success in a number of steps that is a polynomial function of the size of the input. In Shor's algorithm, the alternative possible partitions are associated with a family of nested subspaces. The algorithm works as a randomized algorithm by providing a candidate value for the period in polynomial time, which can be tested in polynomial time.

At first sight, Deutsch's XOR problem is quite different. It involves the determination of a disjunctive property of a Boolean function. But note that determining the period of a function also amounts to determining a disjunctive property of the function: the disjunction over the different subsets s_i of a particular partition of the domain of the function, as opposed to alternative such disjunctions. As we will see below, Deutsch's XOR algorithm works by associating the alternative disjunctions with two Hilbert-space planes that are orthogonal except for an intersection in a ray. From this perspective, the XOR algorithm appears directly as a special case of Simon's algorithm, and all three algorithms can be seen as exploiting the non-Boolean logic represented by the projective geometry of Hilbert space in a similar way.

11.2 Deutsch's XOR algorithm

Deutsch's original XOR algorithm was the first quantum algorithm with a demonstrated speed-up over any classical algorithm performing the same computational task. The algorithm has an even probability of failing, so the improvement in efficiency is achieved only if the algorithm succeeds. A subsequent variation by Cleve [4] avoids this feature, although the speed-up is rather modest: one run of the quantum algorithm versus two runs of a classical algorithm. The Deutsch–Jozsa algorithm [5] achieves an exponential speed-up for a generalized version of the XOR problem known as Deutsch's problem.[1]

[1] But even here, a probabilistic classical algorithm yields a solution with high probability after a few runs – see [11].

In Deutsch's XOR problem [6], a "black box" or oracle computes a Boolean function $f\colon B \to B$, where $B = \{0, 1\}$ is a Boolean algebra (or the additive group of integers mod 2). The problem is to determine whether the function is "constant" (takes the same value for both inputs) or "balanced" (takes a different value for each input). The properties "constant" and "balanced" are two alternative disjunctive properties of the function f (for "constant," $0 \to 0$ and $1 \to 0$ *or* $0 \to 1$ and $1 \to 1$; for "balanced," $0 \to 0$ and $1 \to 1$ *or* $0 \to 1$ and $1 \to 0$). Classically, a solution requires two queries to the oracle, for the input values 0 and 1, and a comparison of the outputs.

Deutsch's algorithm begins by initializing 1-qubit input and output registers to the state $|0\rangle|0\rangle$ in a standard basis (the computational basis). A Hadamard transformation

$$|0\rangle \to |0\rangle + |1\rangle, \tag{11.1}$$

$$|1\rangle \to |0\rangle - |1\rangle \tag{11.2}$$

is applied to the input register (yielding a linear superposition of states corresponding to the two possible input values 0 and 1) followed by a unitary transformation $U_f\colon |x\rangle|y\rangle \to |x\rangle|y \oplus f(x)\rangle$ applied to both registers that implements the Boolean function f:

$$|0\rangle|0\rangle \xrightarrow{H} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|0\rangle, \tag{11.3}$$

$$\xrightarrow{U_f} \frac{1}{\sqrt{2}}(|0\rangle|f(0)\rangle + |1\rangle|f(1)\rangle). \tag{11.4}$$

The final composite state of both registers is then one of two orthogonal states, either (constant)

$$|c_1\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|0\rangle), \tag{11.5}$$

$$|c_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|1\rangle) \tag{11.6}$$

or (balanced)

$$|b_1\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle), \tag{11.7}$$

$$|b_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle). \tag{11.8}$$

Note that the constant states $|c_1\rangle$, $|c_2\rangle$ are not orthogonal to the balanced states $|b_1\rangle$, $|b_1\rangle$. So it is not possible to determine whether the function is constant or balanced by a measurement that identifies, with certainty, the output state as a constant state or a balanced state. The states $|c_1\rangle$, $|c_2\rangle$ and $|b_1\rangle$, $|b_2\rangle$ span two planes

in $\mathcal{H}^2 \otimes \mathcal{H}^2$, the constant plane P_c and the balanced plane P_b, represented by the projection operators

$$P_c = P_{|c_1\rangle} \vee P_{|c_2\rangle}, \tag{11.9}$$

$$P_b = P_{|b_1\rangle} \vee P_{|b_2\rangle}. \tag{11.10}$$

Although the states $|c_1\rangle$, $|c_2\rangle$ are not orthogonal to the states $|b_1\rangle$, $|b_2\rangle$, the planes – which represent quantum disjunctions[2] – are orthogonal, except for an intersection, so their projection operators commute. The intersection is the line (ray) spanned by the vector

$$\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) = \frac{1}{\sqrt{2}}(|c_1\rangle + |c_2\rangle) = \frac{1}{\sqrt{2}}(|b_1\rangle + |b_2\rangle). \tag{11.11}$$

In the prime basis spanned by the states $|0'\rangle = H|0\rangle$, $|1'\rangle = H|1\rangle$, the intersection is the state $|0'\rangle|0'\rangle$, the constant plane is spanned by

$$|0'\rangle|0'\rangle = \frac{1}{\sqrt{2}}(|c_1\rangle + |c_2\rangle), \tag{11.12}$$

$$|0'\rangle|1'\rangle = \frac{1}{\sqrt{2}}(|c_1\rangle - |c_2\rangle) \tag{11.13}$$

and the balanced plane is spanned by

$$|0'\rangle|0'\rangle = \frac{1}{\sqrt{2}}(|b_1\rangle + |b_2\rangle), \tag{11.14}$$

$$|1'\rangle|1'\rangle = \frac{1}{\sqrt{2}}(|b_1\rangle - |b_2\rangle) \tag{11.15}$$

i.e.,

$$P_c = P_{|0'\rangle|0'\rangle} \vee P_{|0'\rangle|1'\rangle}, \tag{11.16}$$

$$P_b = P_{|0'\rangle|0'\rangle} \vee P_{|1'\rangle|1'\rangle}. \tag{11.17}$$

So to decide whether the function f is constant or balanced we could measure the observable with eigenstates $|0'0'\rangle$, $|0'1'\rangle$, $|1'0'\rangle$, $|1'1'\rangle$ on the final state, which is in the three-dimensional subspace orthogonal to the vector $|1'0'\rangle$, either in the constant plane or in the balanced plane. If the state is in the constant plane, we will obtain either the outcome $0'0'$ with probability 1/2 (since the final state is at an angle $\pi/4$ to $|0'0'\rangle$), in which case the computation is inconclusive, or the outcome $0'1'$ with probability 1/2. If the state is in the balanced plane, we will obtain either

[2] Since $P_{|c_1\rangle}$ and $P_{|c_2\rangle}$ are orthogonal, $P_c = P_{|c_1\rangle} \vee P_{|c_2\rangle} = P_{|c_1\rangle} + P_{|c_2\rangle}$, where "$\vee$" represents quantum disjunction: the infimum or span (the smallest subspace containing the two component subspaces). Similarly for P_b.

the outcome $0'0'$ with probability 1/2, in which case the computation is inconclusive, or the outcome $1'1'$ with probability 1/2. So in either case, with probability 1/2, we can distinguish whether the function is constant or balanced in one run of the algorithm by distinguishing between the constant and balanced planes, without evaluating the function at any of its inputs (i.e., without determining in the constant case whether f maps 0 to 0 and 1 to 0, or whether f maps 0 to 1 and 1 to 1, and similarly in the balanced case).[3]

Now, in evaluating the efficiency of the algorithm, we would have to take account of the fact that a computation is required to determine the prime basis, and we would have to count the number of computational steps required for this computation. We would also have to consider whether the measurement could be implemented efficiently. These sorts of issues might be trivial for the XOR algorithm, but in general will not be. To avoid this sort of problem, the number of relevant computational steps in a quantum algorithm is conventionally counted as the number of applications of unitary transformations and measurements required for the successful implementation of the algorithm, where the unitary transformations belong to a standard set of elementary unitary gates that form a universal set, and the measurements are in the computational basis.

Notice that a Hadamard transformation to the final state amounts to dropping the primes in the representation (11.16), (11.17) for the constant and balanced planes (since $H^2 = I$, so $|0'0'\rangle \xrightarrow{H} |00\rangle$, etc.). More precisely, the relationship between the states $|c_1\rangle, |c_2\rangle, |b_1\rangle, |b_2\rangle$ in (11.5)–(11.8) and the constant and balanced planes $P_c = P_{|0'\rangle|0'\rangle} + P_{|0'\rangle|1'\rangle}$ and $P_b = P_{|0'\rangle|0'\rangle} + P_{|1'\rangle|1'\rangle}$ is the same, after the Hadamard transformation of the state, as the relationship between the states $H|c_1\rangle, H|c_2\rangle, H|b_1\rangle, H|b_2\rangle$ and the planes defined by $P_c = P_{|0\rangle|0\rangle} + P_{|0\rangle|1\rangle}$ and $P_b = P_{|0\rangle|0\rangle} + P_{|1\rangle|1\rangle}$. That is, under the Hadamard transformation,

$$|c_1\rangle \to \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |0\rangle|1\rangle), \tag{11.18}$$

$$|c_2\rangle \to \frac{1}{\sqrt{2}}(|0\rangle|0\rangle - |0\rangle|1\rangle) \tag{11.19}$$

and

$$|b_1\rangle \to \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle), \tag{11.20}$$

$$|b_2\rangle \to \frac{1}{\sqrt{2}}(|0\rangle|0\rangle - |1\rangle|1\rangle). \tag{11.21}$$

[3] Equivalently, we could measure the output register. If the outcome is $0'$, the computation is inconclusive. If the outcome is $1'$, we measure the input register. The outcome $1'$ or $0'$ then distinguishes whether the function is constant or balanced.

So the transformed constant plane HP_c is spanned by

$$|0\rangle|0\rangle = \frac{1}{\sqrt{2}}(H|c_1\rangle + H|c_2\rangle), \tag{11.22}$$

$$|0\rangle|1\rangle = \frac{1}{\sqrt{2}}(H|c_1\rangle - H|c_2\rangle) \tag{11.23}$$

and the transformed balanced plane HP_b is spanned by

$$|0\rangle|0\rangle = \frac{1}{\sqrt{2}}(H|b_1\rangle + H|b_2\rangle), \tag{11.24}$$

$$|1\rangle|1\rangle = \frac{1}{\sqrt{2}}(H|b_1\rangle - H|b_2\rangle). \tag{11.25}$$

This allows the transformed constant and balanced planes to be distinguished (with probability 1/2) by a measurement in the computational basis. The outcome 00, when the final state is projected by the measurement onto $|0\rangle|0\rangle$ is inconclusive. The outcome 01, when the final state is projected onto $|0\rangle|1\rangle$, reveals that the final state after the Hadamard transformation was in the transformed constant plane HP_c. It follows that the final state *before* the Hadamard transformation was in the constant plane P_c, and hence that the function f is constant. The outcome 11, when the final state is projected onto $|1\rangle|1\rangle$, reveals that the final state after the Hadamard transformation was in the transformed balanced plane HP_b, and hence by similar reasoning that the function f is balanced. Now, to evaluate the efficiency of measurement, we need only count the number of elementary unitary gates required to implement the Hadamard transformation.

In Cleve's variation, the two registers are initialized to $|0\rangle$ and $|1\rangle$, respectively (instead of to $|0\rangle$ and $|0\rangle$). It is easy to see that, instead of the final state of the two registers ending up as one of two orthogonal states in the constant plane, or as one of two orthogonal states in the balanced plane, the final state now ends up as $\pm|0'1'\rangle$ in the constant plane, or as $\pm|1'1'\rangle$ in the balanced plane, and these states can be distinguished because they are orthogonal. We can distinguish these two possibilities by simply measuring the input register in the prime basis, but since a final Hadamard transformation on the state of the input register takes $|0'\rangle$ to $|0\rangle$ and $|1'\rangle$ to $|1\rangle$), we can distinguish the two planes by measuring the input register in the computational basis. So we can decide with certainty whether the function is constant or balanced after only one run of the algorithm.

Deutsch's XOR problem can be generalized to the problem ("Deutsch's problem") of determining whether a Boolean function $f\colon\ B^n \to B$ is constant or whether it is balanced, where it is promised that the function is either constant or balanced. "Balanced" here means that the function takes the values 0 and 1 an

equal number of times, i.e., 2^{n-1} times each. By exploiting the Cleve variation of the XOR algorithm, the Deutsch–Jozsa algorithm [5] determines whether f is constant or balanced in one run. Since the initial and final Hadamard transformations can be implemented efficiently, i.e., with a number of elementary unitary gates that is only a polynomial function of the size of the input, the algorithm is exponentially faster than any classical algorithm. For a discussion of the Deutsch–Jozsa algorithm from the quantum-logical perspective adopted here, see [3].

11.3 Period-finding algorithms

Simon's problem is to find the period r of a periodic Boolean function $f\colon B^n \to B^n$, i.e., a function for which

$$f(x_i) = f(x_j) \text{ if and only if } x_j = x_i \oplus r, \text{ for all } x_i, x_j \in B^n. \tag{11.26}$$

Note that, since $x \oplus r \oplus r = x$, the function is 2-to-1.

Since f is periodic, the possible outputs of f – the values of f for the different inputs – partition the set of input values into mutually exclusive and collectively exhaustive subsets, and these subsets depend on the period. So, determining the period of f amounts to distinguishing the partition corresponding to the period from alternative partitions corresponding to alternative possible periods. Simon's algorithm solves the problem efficiently, with an exponential speed-up over any classical algorithm (see [14, 15]).

Shor's factorization algorithm [12, 13] exploits the fact that the two prime factors p, q of a positive integer $N = pq$ can be found by determining the period of a function $f(x) = a^x \bmod N$, for any $a < N$ that is coprime to N, i.e., has no common factors with N (other than 1). The period r of $f(x)$ depends on a and N. Once we know the period, we can factor N if r is even and $a^{r/2} \neq -1 \bmod N$, which will be the case with probability greater than 1/2 if a is chosen randomly. (If not, we choose another value of a.) The factors of N are the greatest common factors of $a^{r/2} \pm 1$ and N, which can be found in polynomial time by the Euclidean algorithm. (For these number-theoretic results, see [11, Appendix 4].) So the problem of factorizing a composite integer N that is the product of two primes reduces to the problem of finding the period of a certain function $f\colon Z_s \to Z_n$, where Z_n is the additive group of integers mod n (rather than B^n, the n-fold Cartesian product of a Boolean algebra B, as in Simon's algorithm). Note that $f(x + r) = f(x)$ if $x + r \leq s$. The function f is periodic if r divides s exactly, otherwise it is almost periodic. This adds a complication, but does not change anything essential in the analysis. Shor's algorithm is exponentially faster than any known classical algorithm.

I begin with Simon's algorithm. Here is a sketch of the usual formulation. The input and output registers are initialized to the state $|0\rangle|0\rangle$ in the computational basis (where $|0\rangle$ is an abbreviation for $|0\rangle \ldots |0\rangle = |0 \ldots 0\rangle$) and the state is evolved as follows:

$$|0\rangle|0\rangle \xrightarrow{H} \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle|0\rangle, \tag{11.27}$$

$$\xrightarrow{U_f} \frac{1}{\sqrt{2^n}} \sum_{x} |x\rangle|f(x)\rangle \tag{11.28}$$

$$= \frac{1}{\sqrt{2^{n-1}}} \sum_{x_i} \frac{|x_i\rangle + |x_i \oplus r\rangle}{\sqrt{2}} |f(x_i)\rangle, \tag{11.29}$$

where U_f is the unitary transformation implementing the Boolean function as

$$U_f\colon |x\rangle|y\rangle \to |x\rangle|y \oplus f(x)\rangle. \tag{11.30}$$

A measurement of the output register would leave the input register in a state of the form[4]

$$\frac{|x_i\rangle + |x_i \oplus r\rangle}{\sqrt{2}}. \tag{11.31}$$

This state contains the information r, but summed with an unwanted randomly chosen offset x_i that depends on the measurement outcome. A direct measurement of the state label would yield any $x \in B^n$ equiprobably, providing no information about r.

The application of a final Hadamard transformation yields

$$\frac{|x_i\rangle + |x_i \oplus r\rangle}{\sqrt{2}} \xrightarrow{H} \frac{1}{\sqrt{2^n}} \sum_{y \in B^n} \frac{(-1)^{x_i \cdot y} + (-1)^{(x_i \oplus r)\cdot y}}{\sqrt{2}} |y\rangle \tag{11.32}$$

$$= \sum_{y : r \cdot y = 0} \frac{(-1)^{x_i \cdot y}}{\sqrt{2^{n-1}}} |y\rangle \tag{11.33}$$

where the last equality follows because terms interfere destructively if $r \cdot y = 1$. A measurement of the input register in the computational basis yields a value y (equiprobably) such that $r \cdot y = 0$. Repeating the algorithm sufficiently many times yields enough values y_i so that r can be determined by solving the linear equations $r \cdot y_1 = 0, \ldots, r \cdot y_k = 0$.

To see how the algorithm works quantum logically in terms of the subspaces representing the relevant quantum propositions, consider the case $n = 2$. There

[4] Considering a measurement of the output register here is simply a pedagogical device, for clarity. Instead, we could refer to the reduced state of the input register, which is a mixture of states of the form (11.31). No actual measurement of the output register is required, only a measurement of the input register.

are $2^2 - 1 = 3$ possible values of the period r: 01, 10, 11, and the corresponding partitions are

$$\begin{aligned} r = 01&: \ \{00, 01\}, \{10, 11\}, \\ r = 10&: \ \{00, 10\}, \{01, 11\}, \\ r = 11&: \ \{00, 11\}, \{01, 10\}. \end{aligned}$$

The corresponding states of the input and output registers after the unitary transformation U_f are

$$\begin{aligned} r = 01&: \ \tfrac{1}{2}(|00\rangle + |01\rangle)|f(00)\rangle + \tfrac{1}{2}(|10\rangle + |11\rangle)|f(10)\rangle, \\ r = 10&: \ \tfrac{1}{2}(|00\rangle + |10\rangle)|f(00)\rangle + \tfrac{1}{2}(|01\rangle + |11\rangle)|f(01)\rangle, \\ r = 11&: \ \tfrac{1}{2}(|00\rangle + |11\rangle)|f(00)\rangle + \tfrac{1}{2}(|01\rangle + |10\rangle)|f(01)\rangle. \end{aligned}$$

Notice that this case reduces to the same construction as in Deutsch's XOR algorithm. For $r = 10$ the input register states are

$$|c_1\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |10\rangle), \tag{11.34}$$

$$|c_2\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |11\rangle) \tag{11.35}$$

and for $r = 11$ the input register states are

$$|b_1\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \tag{11.36}$$

$$|b_2\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \tag{11.37}$$

depending on the outcome of a measurement of the output register. Here the orthogonal states $|c_1\rangle, |c_2\rangle$ represent the two subsets of the partition associated with the period $r = 10$, the orthogonal states $|b_1\rangle, |b_2\rangle$ represent the two subsets of the partition associated with the period $r = 11$, and the orthogonal states $|00\rangle + |01\rangle, |10\rangle + |11\rangle$ represent the two subsets of the partition associated with the period $r = 01$.

The three partitions associated with the three possible periods are represented by three planes in $\mathcal{H}^2 \otimes \mathcal{H}^2$, which correspond to the constant and balanced planes in Deutsch's XOR algorithm, and a third orthogonal plane. While the states representing subsets of different partitions associated with different periods are non-orthogonal, the three planes (spanned by these states) are mutually orthogonal, except for an intersection in the ray spanned by the vector $|0'0'\rangle$ in the prime basis (i.e., their projection operators commute):

$r = 01$: plane spanned by $|0'0'\rangle, |1'0'\rangle$,
$r = 10$: plane spanned by $|0'0'\rangle, |0'1'\rangle$ (corresponds to "constant" plane),
$r = 11$: plane spanned by $|0'0'\rangle, |1'1'\rangle$ (corresponds to "balanced" plane).

We cannot identify the period by a measurement that identifies the state of the input register as a state representing a particular subset of a particular partition, because the states representing subsets of different partitions associated with different periods are non-orthogonal. We could identify the plane corresponding to the period by measuring the input register in the prime basis, but – as in Deutsch's XOR algorithm – the final Hadamard transformation (which, as we have seen, amounts to dropping the primes: $|0'0'\rangle \stackrel{H}{\rightarrow} |00\rangle$, etc.) allows the plane corresponding to the period to be identified by a measurement in the computational basis. The three possible periods can therefore be distinguished by measuring the observable with eigenstates $|00\rangle$, $|01\rangle$, $|10\rangle$, $|11\rangle$, except when the state of the register is projected by the measurement onto the state $|00\rangle$ (which occurs with probability 1/2). So the algorithm will generally have to be repeated until we find an outcome that is not 00.

The $n=2$ case of Simon's algorithm essentially reduces to Deutsch's XOR algorithm. In the $n=3$ case (which suffices to exhibit the general feature of the algorithm) there are $2^3-1=7$ possible periods: 001, 010, 011, 100, 101, 110, 111. For the period $r = 001$, the state of the two registers after the unitary transformation U_f is

$$\frac{1}{2\sqrt{2}}(|000\rangle + |001\rangle)|f(000)\rangle + \frac{1}{2\sqrt{2}}(|010\rangle + |011\rangle)|f(010)\rangle + \frac{1}{2\sqrt{2}}(|100\rangle + |101\rangle)|f(100)\rangle + \frac{1}{2\sqrt{2}}(|110\rangle + |111\rangle)|f(110)\rangle. \quad (11.38)$$

A measurement of the output register would leave the input register in one of four states, depending on the outcome of the measurement:

$$\frac{1}{\sqrt{2}}(|000\rangle + |001\rangle) = \frac{1}{2}(|0'0'0'\rangle + |0'1'0'\rangle + |1'0'0'\rangle + |1'1'0'\rangle),$$
$$\frac{1}{\sqrt{2}}(|010\rangle + |011\rangle) = \frac{1}{2}(|0'0'0'\rangle - |0'1'0'\rangle + |1'0'0'\rangle - |1'1'0'\rangle),$$
$$\frac{1}{\sqrt{2}}(|100\rangle + |101\rangle) = \frac{1}{2}(|0'0'0'\rangle + |0'1'0'\rangle - |1'0'0'\rangle - |1'1'0'\rangle),$$
$$\frac{1}{\sqrt{2}}(|110\rangle + |111\rangle) = \frac{1}{2}(|0'0'0'\rangle - |0'1'0'\rangle - |1'0'0'\rangle + |1'1'0'\rangle).$$

Applying a Hadamard transformation amounts to dropping the primes. So if the period is $r = 001$, the state of the input register ends up in the four-dimensional subspace of $\mathcal{H}^2 \otimes \mathcal{H}^2 \otimes \mathcal{H}^2$ spanned by the vectors $|000\rangle$, $|010\rangle$, $|100\rangle$, $|110\rangle$.

A similar analysis applies to the other six possible periods. The corresponding subspaces are spanned by the following vectors:

$$
\begin{aligned}
r = 001&: \ |000\rangle, |010\rangle, |100\rangle, |110\rangle, \\
r = 010&: \ |000\rangle, |001\rangle, |100\rangle, |101\rangle, \\
r = 011&: \ |000\rangle, |011\rangle, |100\rangle, |111\rangle, \\
r = 100&: \ |000\rangle, |001\rangle, |010\rangle, |011\rangle, \\
r = 101&: \ |000\rangle, |010\rangle, |101\rangle, |111\rangle, \\
r = 110&: \ |000\rangle, |001\rangle, |110\rangle, |111\rangle, \\
r = 111&: \ |000\rangle, |011\rangle, |101\rangle, |110\rangle.
\end{aligned}
$$

These subspaces are orthogonal except for intersections in two-dimensional planes. The period can be found by measuring in the computational basis. Repetitions of the measurement will eventually yield sufficiently many distinct values to determine the subspace containing the final state. In this case, it is clear by examining the above list that two values distinct from 000 suffice to determine the subspace, and these are just the values y_i for which $y_i \cdot r = 0$.

Shor's algorithm involves the following steps. A k-qubit input register (whose states are represented on an s-dimensional Hilbert space $\mathcal{H}^s$, where $s = 2^k$) is initialized to the state $|0\rangle \in \mathcal{H}^s$ and the output register to the state $|0\rangle \in \mathcal{H}^N$. A k-fold Hadamard transformation is applied to the input register, followed by the unitary transformation U_f which implements the function $f(x) = a^x \bmod N$:

$$
|0\rangle|0\rangle \xrightarrow{H} \frac{1}{\sqrt{s}} \sum_{x=0}^{s-1} |x\rangle|0\rangle, \tag{11.39}
$$

$$
\xrightarrow{U_f} \frac{1}{\sqrt{s}} \sum_{x=0}^{s-1} |x\rangle|x + a^x \bmod N\rangle. \tag{11.40}
$$

Suppose r divides s exactly. A measurement of the output register in the computational basis would leave the input register in a state of the form

$$
\frac{1}{\sqrt{s/r}} \sum_{j=0}^{s/r-1} |x_i + jr\rangle. \tag{11.41}
$$

The value x_i is the offset, which depends on the outcome i of the measurement of the output register. The sum is taken over the values of j for which $f(x_i + jr) = i$. Since the state label contains the random offset, a direct measurement of the label yields no information about the period.

A discrete quantum Fourier transform for the integers mod s is now applied to the input register, i.e., a unitary transformation

$$
|x\rangle \xrightarrow{U_{\mathrm{DFT}_s}} \frac{1}{\sqrt{s}} \sum_{y=0}^{s-1} \mathrm{e}^{2\pi \mathrm{i} \frac{xy}{s}} |y\rangle, \quad \text{for } x \in Z_s. \tag{11.42}
$$

Note that a Hadamard transformation is a discrete quantum Fourier transform for the integers mod 2, so this step is analogous to the application of the Hadamard transformation in Simon's algorithm. Under the Fourier transformation, the state of the input register undergoes the transition

$$\frac{1}{\sqrt{s/r}}\sum_{j=0}^{s/r-1}|x_i+jr\rangle \xrightarrow{U_{\mathrm{DFT}_s}} \frac{1}{\sqrt{r}}\sum_{k=0}^{r-1}\mathrm{e}^{2\pi\mathrm{i}\frac{x_i k}{r}}|ks/r\rangle, \tag{11.43}$$

where, similarly to the derivation of (11.33), the amplitudes are non-zero only if y is not a multiple k of s/r (i.e., $\sum_{j=0}^{s/r-1}\mathrm{e}^{2\pi\mathrm{i}\frac{jry}{s}} = s/r$ if $y = ks/r$; $\sum_{j=0}^{s/r-1}\mathrm{e}^{2\pi\mathrm{i}\frac{jry}{s}} = 0$ if $y \neq ks/r$). The effect is to shift the offset into a phase factor and invert the period as a multiple of s/r. A measurement of the input register in the computational basis then yields $c = ks/r$, where the value of k is chosen equiprobably by the measurement of the output register. The algorithm is run a number of times until a value of k coprime to r is obtained. Cancelling out c/s to lowest terms then yields k and r as k/r.

Suppose r does not divide s exactly. Then some of the states in (11.41) will have an additional term. For example, suppose $s = r + d$, where $d < r$. Then d of the states in (11.41) will have an extra term and take the form

$$\frac{1}{\sqrt{s/r}+1}\sum_{j=0}^{s/r}|x_i+jr\rangle. \tag{11.44}$$

After the Fourier transformation, the expression (11.43) will contain additional terms with negligible amplitudes for values of $s \neq k$ ($k = 0, 1, \ldots, r-1$) if s/r is large.

Since the value of r is unknown in advance of applying the algorithm, we do not, of course, recognize when a measurement outcome yields a value of k coprime to r. The idea is to run the algorithm, cancel out c/s to lowest terms to obtain a candidate value for r and hence a candidate factor of N, which can then be tested by division into N. Even when we do obtain a value of k coprime to r, some values of a will yield a period for which the method fails to yield a factor of N, in which case we randomly choose a new value of a and run the algorithm with this value. The point is that all these steps are efficient, i.e., can be performed in polynomial time, and, since only a polynomial number of repetitions is required in order to determine a factor with any given probability $p < 1$, the algorithm is a polynomial-time algorithm, achieving an exponential speed-up over any known classical algorithm.

The analysis of the algorithm from a quantum-logical perspective is similar to the above discussion for Simon's algorithm. The essential difference is that the alternative possible partitions are represented by nested subspaces. See [3] for an

account. The algorithm works as a randomized algorithm, producing a candidate value for the period r and hence a candidate factor of N, which can be tested (in polynomial time) by division into N.

11.4 Conclusion

Simon's algorithm and Shor's algorithm work as period-finding algorithms by encoding alternative partitions of the domain of a function, defined by alternative possible periods, as quantum propositions represented by subspaces in a Hilbert space, which are orthogonal except for overlaps. The subspace corresponding to a particular partition is spanned by orthogonal linear superpositions of states associated with the elements in the (mutually exclusive and collectively exhaustive) subsets of the partition. The period-finding algorithm is designed to produce an entangled state in which such superpositions, representing states of an input register, are correlated with distinct orthogonal states of an output register. The reduced state of the input register is then an equal-weight mixture of states spanning the subspace corresponding to the partition, where each state encodes a subset in the partition as a linear superposition of the elements in the subset. Since the subspaces are represented by commuting projection operators, a measurement of the state of the input register in a certain basis can reveal the subspace containing the state, and hence the period associated with the partition, except when the measurement projects the state onto the overlap region. This measurement basis is unitarily related to the computational basis by a known unitary transformation that can be implemented efficiently, so a measurement in the computational basis after this unitary information will yield the same information. This is the function of the final Hadamard transformation or discrete quantum Fourier transformation, and the possibility of an efficient implementation of this transformation is crucial to the efficiency of the algorithm. By contrast with the classical "fast Fourier transform," it is a remarkable feature of the discrete quantum Fourier transform that it can be implemented efficiently.

The information-processing in Deutsch's XOR algorithm has a similar quantum-logical interpretation in terms of the subspace structure of Hilbert space. The problem here is to distinguish two alternative disjunctive properties of a function ($0 \rightarrow 0$ and $1 \rightarrow 0$ *or* $0 \rightarrow 1$ and $1 \rightarrow 1$ for a constant function, versus $0 \rightarrow 0$ and $1 \rightarrow 1$ *or* $0 \rightarrow 1$ and $1 \rightarrow 0$ for a balanced function), which are encoded as two planes in a four-dimensional Hilbert space (orthogonal except for an overlap). Each disjunct in the disjunction is a conjunction of two elements (e.g., $0 \rightarrow 0$ *and* $1 \rightarrow 0$). The plane corresponding to a particular disjunction is spanned by a pair of states that encode the elements of the conjunctions as linear superpositions. The algorithm is designed to produce one of these states, depending on which

disjunction is true of the function. From this perspective, the XOR algorithm appears directly as a special case of Simon's algorithm.

The first stage of a quantum algorithm designed to evaluate some global property of a function involves the creation of an entangled state of the input and output registers in which every value in the domain of the function is correlated with a corresponding value in its range. This is referred to as "quantum parallelism" and is often cited as the source of the speed-up in a quantum computation. The idea is that a quantum computation is something like a massively parallel classical computation, for all possible values of a function. This appears to be Deutsch's view [7]: in an Everettian many-worlds interpretation of quantum mechanics, the parallel computations can be regarded as taking place in parallel universes. (For a critique, see [16].)

From the quantum-logical perspective outlined here, the picture is entirely different. Rather than "computing all values of a function at once," a quantum algorithm achieves an exponential speed-up over a classical algorithm precisely by avoiding the computation of *any* values of the function at all. This is redundant information for a quantum algorithm but essential information for a classical algorithm. A quantum algorithm works by exploiting the non-Boolean logic represented by the projective geometry of Hilbert space to encode a global property of a function (such as a period, or a disjunctive property) as a subspace in Hilbert space, which can be efficiently distinguished from alternative subspaces, corresponding to alternative global properties, by a measurement (or sequence of measurements) that identifies the target subspace as the subspace containing the final state produced by the algorithm. The point of the procedure is precisely to avoid the evaluation of the function in the determination of the global property, in the sense of producing a value in the range of the function for a value in its domain, and it is this feature – impossible in the Boolean logic of classical computation – that leads to the speed-up relative to classical algorithms.

In a sense, evaluating the efficiency of a quantum algorithm relative to a classical algorithm is like comparing apples and oranges because the information processing is so different – hence the convention of counting the number of steps in a quantum computation as the number of applications of unitary gates from a standard set of elementary unitary gates, and restricting measurements to a standard basis. The efficient implementation of the final Fourier or Hadamard transform is, of course, crucial to the efficiency of the algorithm – if this were not possible, there would be no point in building a quantum computer. Nevertheless, even before the application of the final transformation, the subspace representing the global property of the function already contains the state, i.e., the quantum proposition representing a particular global property, as opposed to alternative possible

global properties, has already been selected as true by the state. It is then simply a question of determining, by a suitable measurement, which of the alternative propositions is true, i.e., which of the alternative propositions is represented by the state.

This is quite different from the claim that the quantum state already contains the information correlating all values in the domain of the function with corresponding values in the range. There is, in principle, no way of extracting this information by any measurement, whereas the information about the global property encoded in a subspace can be extracted by a suitable measurement. While we certainly need to consider the question of the efficiency of this measurement in comparing a quantum computation with a classical computation, there is a real sense in which the source of the exponential speed-up in a quantum computation ultimately turns on the different logical structures of the classical and quantum event spaces.

It is this difference that makes possible the representation of a global or disjunctive property of a function as a subspace containing the quantum state of a computer register in a quantum computation, while a classical computation represents such a property as a subset containing the classical state. The possibility of an exponential speed-up arises because the classical state can end up in a particular subset only by ending up at a particular point in the subset, representing a particular pair of input–output values correlated by the function. So the evolution of the state to that point necessarily involves a computation that keeps track of the information required to exclude all the other points in the subset. By contrast, a quantum state can end up in a particular subspace that represents a global or disjunctive property without representing a particular pair of input–output values correlated by the function, i.e., without keeping track of the information required by a classical computation. If the size of the subset representing the global property in a classical computation grows exponentially with the size of the input, one can see the possibility of an exponential slow-down relative to a quantum computation, assuming the efficient extraction of the information about the global property represented in the final quantum state.

Acknowledgments

Support for research leading to this chapter is acknowledged from the University of Maryland General Research Board (2005), the National Science Foundation (2006), and the Perimeter Institute for Theoretical Physics in Waterloo, Canada, where the text was written during a stay as a long-term visiting researcher in 2006. I thank Richard Jozsa for illuminating correspondence.

References

[1] A. Barenco, A. Ekert, K.-A. Suominen, and P. Törmä, Approximate quantum Fourier transform and decoherence, *Phys. Rev. A* **54**, 139–146 (1996).

[2] A. Barenco, Quantum computation: an introduction, in *Introduction to Quantum Computation and Information*, H.-K. Lo, S. Popescu, and T. Spiller (eds.) (Singapore: World Scientific, 1998), pp. 143–183.

[3] J. Bub, (2006), Quantum computation from a quantum logical perspective, *Quantum Information and Computation* **7**, 281–296 (2007).

[4] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Quantum algorithms revisited, *Proc. Roy. Soc. London A* **454**, 339–354 (1998).

[5] D. Deutsch and R. Jozsa, Rapid solutions of problems by quantum computation, *Proc. Roy. Soc. London A* **439**, 553–558 (1992).

[6] D. Deutsch, Quantum theory, the Church–Turing principle and the universal quantum computer, *Proc. Roy. Soc. London A* **400**, 97–117 (1985).

[7] D. Deutsch, *The Fabric of Reality* (London: Penguin, 1997).

[8] A. Ekert and R. Jozsa, Quantum computation and Shor's factoring algorithm, *Rev. Mod. Phys.* **68**, 733–753 (1996).

[9] R. Jozsa, Quantum algorithms and the Fourier transform, *Proc. Roy. Soc. London A* **454**, 323–337 (1998).

[10] R. Jozsa, Quantum factoring, discrete logarithms and the hidden subgroup problem, *IEEE Computing Sci. Eng.* **3**, 34–43 (2001).

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

[12] P. W. Shor, Algorithms for quantum computation: discrete logarithms and factoring, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (Los Alamitos, CA: IEEE Press, 1994), pp. 124–134.

[13] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM J. Computing* **26**, 1484–1509 (1997).

[14] D. R. Simon, On the power of quantum computation, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (Los Alamitos, CA: IEEE Press, 1994), pp. 116–123.

[15] D. R. Simon, On the power of quantum computation, *SIAM J. Computing* **26**, 1474–1483 (1994).

[16] A. Steane, A quantum computer needs only one universe, *Stud. Hist. Philos. Mod. Phys. B* **34**, 469–478 (2003).

12

Quantum mechanics, quantum computing, and quantum cryptography

Tai Tsun Wu

12.1 Introduction

Quantum mechanics is much more than the superposition of states. Since quantum computing and quantum cryptography both involve quantum effects in essential ways, they are necessarily dealing with quantum systems, which need to be studied on the basis of quantum mechanics. It is the purpose of this chapter to initiate such a study with the spatial variable taken into account [1].

The physics of quantum mechanics is described by the Schrödinger equation. In the simple case of a particle of mass m in an external potential $V(\mathbf{r})$, the time-independent Schrödinger equation is

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r})+V(\mathbf{r})\psi(\mathbf{r})=k^2\psi(\mathbf{r}). \tag{12.1}$$

When there is only one spatial dimension x instead of the usual three, then (12.1) reduces to

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi(x)}{\mathrm{d}x^2}+V(x)\psi(x)=k^2\psi(x). \tag{12.2}$$

It is customary to choose units so that

$$\hbar=1 \qquad \text{and} \qquad 2m=1.$$

With such units, (12.2) is

$$-\frac{\mathrm{d}^2\psi(x)}{\mathrm{d}x^2}+V(x)\psi(x)=k^2\psi(x). \tag{12.3}$$

In this chapter, we shall deal extensively with (12.3). It is the use of the Schrödinger equations (12.1)–(12.3) that distinguishes this chapter from most of the other chapters in this book.

Consider next the second topic in the title of this chapter: quantum computing. This is a deep and rapidly developing topic. As is appropriate for such fields, many

Philosophy of Quantum Information and Entanglement, ed. A. Bokulich and G. Jaeger. Published by Cambridge University Press.

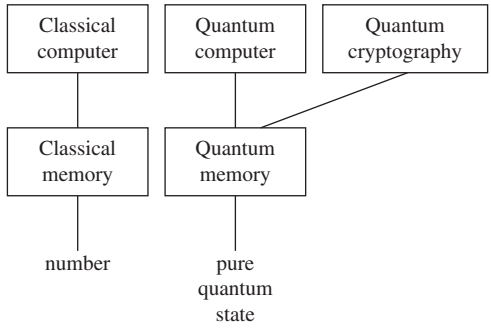


Fig. 12.1 Quantum memory.

of the theoretical studies are based on the simplest possible models [2, 3]. These models typically do not take the spatial variables into account. In this chapter, some aspects of quantum computers are to be treated with the Schrödinger equation (12.3), which involves a spatial variable explicitly.

No quantum computer has ever been built; indeed, none has ever been designed. However, no matter how quantum computers develop in the future, two components are likely to be essential.

A first necessary component is the quantum memory – sometimes called the quantum register. While the content of a classical memory is a number, that of a quantum memory is a pure quantum state. The central role of the quantum memory in the context of quantum computing and quantum cryptography is shown schematically in Figure 12.1. The Schrödinger equation (12.3) is to be used to analyze some simple cases of quantum memory.

A second necessary component of a quantum computer is the connection between the various components. On the basis of the present knowledge, these components are likely to be connected by optical fibers. Although it seems natural to apply quantum electrodynamics to study the properties of such optical fibers, there are many unresolved problems and therefore it is at present premature to treat this topic here. In other words, nothing further will be said in this chapter about such connections between the components.

In summary, the Schrödinger equation is to be used to analyze some simple cases of quantum memories, and the result of this analysis is to be applied to quantum cryptography.

12.2 Quantum memory

In classical computing, the content of a classical memory needs to be altered as the computation proceeds; in quantum computing, the content of a quantum memory is similarly altered. Since the content of a quantum memory is a pure quantum state,

the change from one pure state to another is described by a unitary transform. In other words, mathematically we apply unitary transforms to the content of a quantum memory.

From the engineering point of view, the more important issue is *how* to apply a unitary transform to the content of a quantum memory.

Since a quantum memory is necessarily small, the most natural and practical way to change its content is to send a signal from outside of the memory to interact or scatter from this. This interaction or scattering, being quantum in nature, must be described by the Schrödinger equation, such as (12.1). Equation (12.1) involves spatial coordinates and differentiations with respect to these variables.

Why is the inclusion of spatial variables so important? First, spatial variables are physically present in any case. Secondly, and perhaps more importantly, the presence of spatial variables makes it possible to apply the powerful and well-developed theory of scattering [4] to the study of quantum memory. In scattering theory, it is essential to have at least one spatial variable; otherwise it is not possible to give a meaning to the in and out states [5].

The use of scattering theory in the context of quantum memory has one more major advantage. In addition to the contents of the quantum memory before and after scattering, there are also the incoming and outgoing wavefunctions $\psi^{\text{in}}(\mathbf{r})$ and $\psi^{\text{out}}(\mathbf{r})$ of the signal from outside the memory. These $\psi^{\text{in}}(\mathbf{r})$ and $\psi^{\text{out}}(\mathbf{r})$ have no analog in the treatments without spatial dimensions, and they contain valuable information about the quantum memory. In particular, some of the operations that have been deemed impossible in previous models are found to present no difficulty from the present point of view with spatial dimensions.

One of the major differences between a classical memory and a quantum memory is the following. While, by now, almost all classical memories are digital, any quantum memory, with a pure quantum state as its content, is necessarily non-digital. Therefore, while a classical memory can be read, for example, *exactly*, the reading of a quantum memory necessarily involves errors. A prevalent form of error in the case of quantum memory is decoherence, i.e., its content deviates from being a pure quantum state. An especially simple form of decoherence will be considered in Section 12.5.

Since the content of a quantum memory is a pure quantum state, this must be true both before and after scattering. This condition limits severely what incident waves are allowed to scatter from the quantum memory. This basic concept of an *admissible* incident wave has the following origin [1, 6].

Let $|j\rangle$ be linearly independent quantum states for the memory; a state, which must be a pure state, for the memory is of the form $\sum_j a_j|j\rangle$. In order to perform any operation on the memory, a wave or equivalently a particle, or perhaps several particles, is sent from the outside to interact with the memory. Let $\psi^{\text{in}}(\mathbf{r})$

denote the wavefunction for this incident wave. Then the in field for the scattering process is

$$\Psi^{\text{in}} = \left(\sum_j a_j^{\text{in}}|j\rangle\right)\psi^{\text{in}}(\mathbf{r}). \tag{12.4}$$

Note that $\psi^{\text{in}}(\mathbf{r})$ is our choice to accomplish whatever the purpose of the scattering is.

The corresponding out state is in general a sum of products

$$\Psi^{\text{out}} = \sum_j |j\rangle\psi_j^{\text{out}}(\mathbf{r}). \tag{12.5}$$

This Ψ^{out} contains information about the state of the memory after scattering and the behavior of the scattered wave. Later, the scattered wave as described by $\psi_j^{\text{out}}(\mathbf{r})$ moves away from the memory and the information contained there is no longer available. This means that the final state of the memory is given by Ψ^{out} *averaged* over this information that is no longer available. On the other hand, in order for the quantum memory to function, the final state of the memory must be a pure state. This condition of being a pure state must mean that the various $\psi_j^{\text{out}}(\mathbf{r})$s are related to each other. Intuitively, we expect that Ψ^{out} must also be of the form of (12.4), i.e., (12.5) must reduce to

$$\Psi^{\text{out}} = \left(\sum_j a_j^{\text{out}}|j\rangle\right)\psi^{\text{out}}(\mathbf{r}). \tag{12.6}$$

In other words, not only the in state but also the out state is unentangled. This leads to the important concept of admissible ψ^{in}. The proof of (12.6) can be found in Refs. [1, 6].

A ψ^{in} is defined to be *admissible* if, for all $\sum_j a_j^{\text{in}}|j\rangle$ of interest, the corresponding Ψ^{out} is of the form of (12.6).

When $\psi^{\text{in}}(\mathbf{r})$ is an admissible incident field, then it follows from (12.4) and (12.6) that the initial state of the quantum memory is $\sum_j a_j^{\text{in}}|j\rangle$, while the final state is $\sum_j a_j^{\text{out}}|j\rangle$. In other words, the use of this admissible $\psi^{\text{in}}(\mathbf{r})$ performs a unitary transform on the content of the quantum memory, changing it from $\sum_j a_j^{\text{in}}|j\rangle$ to $\sum_j a_j^{\text{out}}|j\rangle$.

An important phrase in the above definition of an admissible ψ^{in} is "for all $\sum_j a_j^{\text{in}}|j\rangle$ of interest." The reason for the inclusion of this phrase in the definition is that in general it is not known what the content of the quantum memory, $\sum_j a_j^{\text{in}}|j\rangle$, is initially. If (12.6) holds for some $\sum_j a_j^{\text{in}}|j\rangle$ but not for others of interest, then it is necessary to perform the task of determining what the state is in the quantum memory before applying $\psi^{\text{in}}(\mathbf{r})$ to interact with it.

On the other hand, the inclusion of this phrase leads immediately to the following question: is there any admissible ψ^{in} at all? In fact, what is needed is not merely the existence of one or a few admissible ψ^{in}. In order to perform the unitary transforms on the quantum memory to carry out the quantum computations, it is necessary to have a sufficiently large collection of admissible ψ^{in}.

The central question is therefore the following: is there a sufficiently large collection of admissible ψ^{in} so that the necessary operations of quantum computing can be performed through scattering? An alternative way of asking this question is as follows: is the present way of interrogating and controlling the content of a quantum memory through scattering self-consistent or not?

In order to answer this question of self-consistency, it is essential to have an explicit example. Such examples are discussed in the next section. Note that this self-consistency is guaranteed by the construction of a single non-trivial example.

12.3 A potential model of quantum memory

That the out state is unentangled means that the right-hand side of (12.5) always reduces to that of (12.6). This is most easily satisfied if there is only one possible form for $\psi_j^{\text{out}}(\mathbf{r})$. What this means is that, no matter what j is, this $\psi_j^{\text{out}}(\mathbf{r})$ is always proportional to a specific function of $\mathbf{r}$. As an example, consider scattering by a spherical potential. If the incident wave is a plane wave, then the scattered wave is in general complicated, being a superposition of the partial waves. If, however, the in wave contains only one partial wave, then the out wave also contains only this particular partial wave. Therefore, that there is only one possible form for $\psi_j^{\text{out}}(\mathbf{r})$ is closely related to *symmetry*.

In this section, the model for quantum memory is to be sought in the context of potential scattering [1, 6]. Since there are at least two linearly independent quantum states for the memory, it is necessary to consider coupled-channel scattering. With the idea of constructing the simplest model, the number of channels should be chosen to be the smallest, namely two, and the number of spatial dimensions should also be chosen this way, namely one. Thus we are using the one-dimensional Schrödinger equation (12.3), where the wavefunction ψ has two components,

$$\psi(x) = \begin{bmatrix} \psi_1(x) \\ \psi_2(x) \end{bmatrix}, \tag{12.7}$$

and the potential $V(x)$ is a 2×2 Hermitian matrix

$$V(x) = \begin{bmatrix} V_{11}(x) & V_{12}(x) \\ V_{21}(x) & V_{22}(x) \end{bmatrix}. \tag{12.8}$$

If the two channels are not coupled, then (12.3) reduces to two scalar equations and thus cannot describe a non-trivial quantum memory. To avoid this uninteresting case, the potential $V(x)$ of (12.8) is chosen such that it cannot be diagonalized simultaneously for all x.

Thus the scattering problem under consideration, in the time-independent case, deals with the coupled Schrödinger equations [7, 8]

$$\begin{aligned} -\frac{\mathrm{d}^2\psi_1(x)}{\mathrm{d}x^2} + V_{11}(x)\psi_1(x) + V_{12}(x)\psi_2(x) &= k^2\psi_1(x), \\ -\frac{\mathrm{d}^2\psi_2(x)}{\mathrm{d}x^2} + V_{21}(x)\psi_1(x) + V_{22}(x)\psi_2(x) &= k^2\psi_2(x). \end{aligned} \tag{12.9}$$

The one-dimensional analog of a spherical potential is a symmetric potential. Thus, the 2×2 potential of (12.8) is chosen to be an even function of x, i.e.,

$$V(-x) = V(x). \tag{12.10}$$

With condition (12.10), the scattering processes for the even and odd incident waves are independent. For example, consider the even case, in which the Ψ^{in} and Ψ^{out} of (12.4) and (12.6) take the forms

$$\Psi^{\mathrm{in}} = \begin{bmatrix} a_1^{\mathrm{in}} \\ a_2^{\mathrm{in}} \end{bmatrix} \mathrm{e}^{-\mathrm{i}k|x|} \qquad \text{and} \qquad \Psi^{\mathrm{out}} = \begin{bmatrix} a_1^{\mathrm{out}} \\ a_2^{\mathrm{out}} \end{bmatrix} \mathrm{e}^{\mathrm{i}k|x|}. \tag{12.11}$$

Thus, the 2×2 scattering matrix $S_+(k)$ is defined by

$$\begin{bmatrix} a_1^{\mathrm{out}} \\ a_2^{\mathrm{out}} \end{bmatrix} = S_+(k) \begin{bmatrix} a_1^{\mathrm{in}} \\ a_2^{\mathrm{in}} \end{bmatrix}. \tag{12.12}$$

The scattering matrix $S_-(k)$ for the odd case is defined similarly. This shows that, in this potential model, there are many admissible ψ^{in}.

It is instructive to have an explicit example where the existence of this S is verified directly. Such an example is provided by using the simplest potential in the form of a Fermi pseudo-potential. This has been carried out in great detail in Ref. [1]. See Section 12.7 below.

Thus the two-channel Schrödinger equations (12.9) together with (12.8) and (12.10) give a simple potential model for the quantum memory.

12.4 Write, read, and reset

Returning to the general case, consider an admissible ψ^{in} so that (12.4) and (12.6) both hold. This means that the unitary transform that changes the pure quantum state $\sum_j a_j^{\mathrm{in}}|j\rangle$ in the memory to the pure quantum state $\sum_j a_j^{\mathrm{out}}|j\rangle$ can be carried out using the admissible ψ^{in}. This ψ^{in} is designated by

$$\psi^{\text{in}}\left(\sum_j a_j^{\text{in}}|j\rangle \to \sum_j a_j^{\text{out}}|j\rangle\right). \tag{12.13}$$

Let

$$|X\rangle = \sum_j x_j|j\rangle \qquad \text{and} \qquad |Y\rangle = \sum_j y_j|j\rangle \tag{12.14}$$

be two possible pure states of a quantum memory. In general, there is no $\psi^{\text{in}}(|X\rangle \to |Y\rangle)$, i.e., there is no ψ^{in} that can be scattered from the quantum memory to carry out the unitary transform from $|X\rangle$ to $|Y\rangle$. This can be seen most easily by counting the number of continuous parameters. Let n be the number of linearly independent states in a quantum memory; then the number of real parameters for the unitary transform is n^2. If the number of independent real parameters for all admissible ψ^{in} is less than n^2, then there are not enough admissible ψ^{in} to give $\psi^{\text{in}}(|X\rangle \to |Y\rangle)$ for all $|X\rangle$ and $|Y\rangle$. As a specific example, for the case of Section 12.3 $n = 2$ but there is only one real continuous parameter for the admissible ψ^{in}.

Therefore, the unitary transform from $|X\rangle$ to $|Y\rangle$ is to be accomplished by a sequence of admissible ψ^{in}. Let ϕ^{in} denote such a finite sequence:

$$\phi^{\text{in}} = \{\psi^{\text{in}(1)}, \psi^{\text{in}(2)}, \psi^{\text{in}(3)}, \ldots, \psi^{\text{in}(N)}\}. \tag{12.15}$$

The notation of (12.13) is generalized to

$$\phi^{\text{in}}(|X\rangle \to |Y\rangle) \tag{12.16}$$

to mean that the sequence (12.15) of admissible incident waves can be used to accomplish the unitary transform from the state $|X\rangle$ to the state $|Y\rangle$. More precisely, (12.16) means that there is a sequence of pure states $|X_i\rangle$, $i = 0, 1, 2, \ldots, N$, such that

$$|X_0\rangle = |X\rangle, \qquad |X_N\rangle = |Y\rangle,$$

and

$$\psi^{\text{in}(j)}(|X_{j-1}\rangle \to |X_j\rangle) \tag{12.17}$$

in the sense of (12.13) for $j = 1, 2, \ldots, N$.

From here on, it will be assumed that, for any pair of quantum states $|X\rangle$ and $|Y\rangle$ in the memory, there is a ϕ^{in} of a finite sequence of admissible incident waves such that $\phi^{\text{in}}(|X\rangle \to |Y\rangle)$. The operations of writing, reading, and resetting a quantum memory are to be discussed here under this assumption.

For a *classical memory*, the most basic operations that can be performed are write, read, and reset. More precisely, in this case, there is a standard number, usually chosen to be 0. *Writing* means changing the content of the classical memory

from this 0 to whatever number we want to put into it. *Reading* means determining what number is in the classical memory. *Resetting* means replacing the content of the classical memory by 0.

For a *quantum* memory, these three operations have the following analogous meanings. One particular pure quantum state is chosen to be the standard state, to be designated as s. *Writing* in this case means changing the content of a quantum memory from this s to a desired given superposition of quantum states. *Reading* means determining what superposition of quantum states has been written there. *Resetting* means replacing the content of the quantum memory by the standard state s.

More precisely, these three operations may be described as follows.

Writing. The issue here is how a given superposition of quantum states can be written on a quantum memory after the memory has been reset to s.

Reading. In this case the issue is how to extract from the memory whatever superposition of states has been written there.

Resetting. For a quantum memory in an arbitrary superposition of quantum states, it is desired to erase this information and replace it by s. This should be done without violating unitarity and time-reversal invariance. It may be noted that, while time reversal is not an invariance of nature, its violation involves the second and third generations of quarks [9–12], which are not expected to play any significant role in quantum memory.

These desired operations are to be accomplished through scattering as follows.

(a) *Writing*. The simplest of these three operations is writing. Since the quantum memory has been reset beforehand, let $\sum_j a_j^{(0)}|j\rangle$ be the given quantum state to be written on the memory. Then, writing can be accomplished by using $\phi^{\text{in}}\big(s \to \sum_j a_j^{(0)}|j\rangle\big)$.

(b) *Reading*. Reading from a quantum memory is very different from, and more complicated than, writing on such a memory. Consider a quantum memory in a state $\sum_j a_j^{\text{in}}|j\rangle$. It is desired to determine the values of these coefficients a_j^{in} by interrogating this memory. More precisely, using the sequence of admissible ψ^{in}s as given by (12.15), consider the successive scattering processes:

$$\begin{aligned}
& \textstyle\sum_j a_j^{\text{in}(1)}|j\rangle\psi^{\text{in}(1)}(\mathbf{r}) \to \sum_j a_j^{\text{out}(1)}|j\rangle\psi^{\text{out}(1)}(\mathbf{r}) \to a_j^{\text{out}(1)} = a_j^{\text{in}(2)} \\
\to\ & \textstyle\sum_j a_j^{\text{in}(2)}|j\rangle\psi^{\text{in}(2)}(\mathbf{r}) \to \sum_j a_j^{\text{out}(2)}|j\rangle\psi^{\text{out}(2)}(\mathbf{r}) \to a_j^{\text{out}(2)} = a_j^{\text{in}(3)} \\
\to\ & \textstyle\sum_j a_j^{\text{in}(3)}|j\rangle\psi^{\text{in}(3)}(\mathbf{r}) \to \sum_j a_j^{\text{out}(3)}|j\rangle\psi^{\text{out}(3)}(\mathbf{r}) \to a_j^{\text{out}(3)} = a_j^{\text{in}(4)} \\
\to\ & \cdots \\
\to\ & \textstyle\sum_j a_j^{\text{in}(N)}|j\rangle\psi^{\text{in}(N)}(\mathbf{r}) \to \sum_j a_j^{\text{out}(N)}|j\rangle\psi^{\text{out}(N)}(\mathbf{r}). \qquad (12.18)
\end{aligned}$$

The meanings of the arrows in (12.18) are as follow. Similarly to in (12.4), the quantity in the upper left corner is the in field for the first scattering process. The first arrow on the first line means that the in field to the left of the arrow leads to the corresponding out state to the right of the arrow, namely, $\sum_j a_j^{\text{out}(1)}|j\rangle\psi^{\text{out}(1)}(\mathbf{r})$. This is the first scattering. As indicated by the second arrow on the first line, after this first scattering the coefficients $a_j^{\text{out}(1)}$ of the quantum state in the memory describe the initial state $\sum_j a_j^{\text{in}(2)}|j\rangle$ of the memory for the second scattering. Thus each line of (12.18) describes one scattering process, and there are altogether N successive scatterings.

It is seen from (12.18) that, corresponding to the list (12.15), we have the list of $\psi^{\text{out}}(\mathbf{r})$s, i.e.,

$$\psi^{\text{out}(1)}(\mathbf{r}),\ \psi^{\text{out}(2)}(\mathbf{r}),\ \psi^{\text{out}(3)}(\mathbf{r}),\ \ldots,\ \psi^{\text{out}(N)}(\mathbf{r}). \tag{12.19}$$

From the quantities listed in (12.15) and (12.19) together with their interference, the values of a_j^{in} are obtained. Once the a_j^{in} are found, then the $a_j^{\text{out}(N)}$ can be calculated. An additional scattering using any one of the admissible $\phi^{\text{in}}\big(\sum_j a_j^{\text{out}(N)}|j\rangle \to \sum_j a_j^{\text{in}}|j\rangle\big)$ returns the quantum memory to its initial state.

(c) *Resetting*. With reading as prescribed above, resetting is now straightforward. Suppose we want to reset a quantum memory in the initial state $\sum_j a_j^{\text{in}}|j\rangle$ to the standard state s. This resetting consists of the following two steps. (i) Read the memory as described above. (Note that, after this process of reading has been performed, the memory is in the original initial state $\sum_j a_j^{\text{in}}|j\rangle$.) (ii) Apply an additional scattering using $\phi^{\text{in}}(\sum_j a_j^{\text{in}}|j\rangle \to s)$. This leaves the memory in the desired standard state s.

This completes the description of the quantum memory together with writing, reading, and resetting, all performed through scattering from the memory.

These three operations of writing, reading, and resetting depend crucially on each other: writing requires the ability to reset, reading that of writing, and resetting that of reading.

The importance of interference cannot be overemphasized. In the description of the scattering process, there are not only the in field Ψ^{in} of (12.4) and the out field Ψ^{out} of (12.6), but also the total field Ψ. While Ψ^{in} and Ψ^{out} are unentangled, Ψ is not. When the spatial coordinates are neglected, there is no analog to Ψ.

Interference can occur in different ways. For example, the two fields interfering with each other can be going in opposite directions or the same direction. In the case of one spatial dimension, these two cases are described respectively by

$$f_1(x) = Ae^{ikx} + Be^{-ikx}, \tag{12.20}$$

and

$$f_2(x) = Ae^{ikx} + Be^{ikx}, \tag{12.21}$$

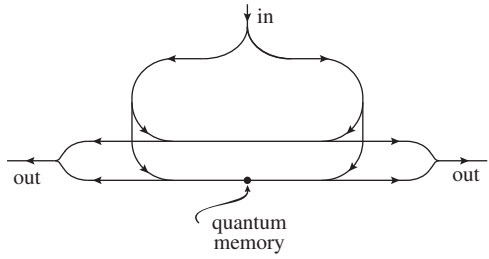


Fig. 12.2 A schematic diagram for the use of interference of waves going in the same direction.

where A and B may be taken to be column vectors. Thus,

$$f_1^\dagger(x)\, f_1(x) = A^\dagger A + B^\dagger B + 2\,\mathrm{Re}(B^\dagger A e^{2ikx}) \tag{12.22}$$

and

$$f_2^\dagger(x)\, f_2(x) = A^\dagger A + B^\dagger B + 2\,\mathrm{Re}(B^\dagger A). \tag{12.23}$$

In both cases, interference gives information about $B^\dagger A$.

Interference has been discussed in Ref. [1] on the basis of (12.20). A schematic diagram for using (12.21) instead is shown in Figure 12.2.

12.5 Decoherence due to the finite length of the pulse

All the developments so far are based on the time-independent Schrödinger equation. In this sense, time has not been introduced. On the other hand, in order to obtain a unitary transform that changes the content of a quantum memory from $|X\rangle$ to $|Y\rangle$, a sequence (12.15) of admissible incident waves is used, as seen explicitly from (12.17). Physically, this means that these admissible incident waves must scatter from the quantum memory sequentially in time.

When a segment in time of an admissible incident wave is used to scatter from a quantum memory, the out field is no longer of the form of (12.6), i.e., there is unavoidably some entanglement between the quantum state in the memory and the ψ^{out}. In other words, there is decoherence in the quantum memory, meaning deviation from a pure quantum state [6, 13]. The shorter this segment in time is, the larger the decoherence. On the other hand, the longer these incident fields in time are, the slower the quantum memory operates. Such slow-down in the quantum memory also reduces the speed of any system that makes essential use of quantum memories, including quantum computing and quantum communication. A compromise is therefore necessary.

In the potential model of Section 12.3, the underlying equations are given by (12.9), which are time-independent. In order to study the scattering by a pulse, the

k^2 in (12.9) must be replaced by a suitable time derivative. There are two ways to do this, namely,

$$k^2 \to \mathrm{i}\,\frac{\partial}{\partial t}, \tag{12.24}$$

or

$$k \to \mathrm{i}\,\frac{\partial}{\partial t}. \tag{12.25}$$

These two replacements lead respectively to the non-relativistic

$$\left[-\frac{\partial^2}{\partial x^2} + V(x)\right]\psi(x,t) = \mathrm{i}\,\frac{\partial}{\partial t}\,\psi(x,t) \tag{12.26}$$

and the relativistic Schrödinger equation

$$\left[-\frac{\partial^2}{\partial x^2} + V(x)\right]\psi(x,t) = -\frac{\partial^2}{\partial t^2}\,\psi(x,t). \tag{12.27}$$

In both cases, the complex wavefunction $\psi(x,t)$ is

$$\psi(x,t) = \begin{bmatrix} \psi_1(x,t) \\ \psi_2(x,t) \end{bmatrix}. \tag{12.28}$$

The question is which one of these two replacements should be used in studying quantum memory in general and the decoherence due to the finite length of the pulse in particular?

The answer is that the replacement (12.25) should be used, not (12.24). There are several reasons for this answer, and some of them are as follows.

(1) Nature is relativistic. Although non-relativistic approximations are useful under many circumstances, they have to be avoided when they cause unnecessary complications.
(2) More specifically, there is a practical reason for quantum memory in general. While there are, in principle, various choices for the wave or particle used to scatter from the memory, the most practical one at present is to use the electromagnetic wave or photon. Since the photon is massless, it is never non-relativistic.
(3) If (12.24) is used, the propagation on the basis of (12.26), even in the absence of a scatterer, widens the individual pulses without limit. This has the consequence that the pulses merge together.
(4) In order to understand decoherence within the present context, it is essential to give a meaning to "the duration of the incoming wave." While this term does not have a unique interpretation, the "duration" should not vary greatly as a function of time. This is automatically satisfied if the relativistic Schrödinger equation (12.27) is used. If one insists on using the non-relativistic Schrödinger equation (12.26), then it is necessary to introduce an additional parameter, which limits the time interval when the waves are allowed to propagate. With this maximum allowed time interval, the spreading of any incident wave of a finite duration can be controlled. The price to be paid is that the desired solutions then depend on this new parameter.

Consider $x < 0$ for definiteness. Then the $e^{-ik|x|}$ that appears in the Ψ^{in} of (12.11) is e^{ikx}. With the desired replacement (12.25), this e^{ikx} becomes

$$e^{ikx} \to e^{ikx}e^{-ikt} = e^{-ik(t-x)}. \tag{12.29}$$

This is an infinitely long progressive wave. In order to get a pulse of finite length T, this in field is further replaced by

$$e^{-ik(t-x)} \to e^{-ik_0(t-x)}g(t-x), \tag{12.30}$$

where the scalar function $g(t)$ has the property

$$g(t) = 0, \quad \text{for } |t| \geq \tfrac{1}{2}T. \tag{12.31}$$

In (12.30), k_0 is the central frequency. When the right-hand side of (12.30) is used with a $g(t)$ that is not identically zero, the content of the quantum memory is no longer a pure state.

Let $g(t)$ be normalized by

$$\int_{-T/2}^{T/2} |g(t)|^2 \, dt = T. \tag{12.32}$$

Then it is a well-defined problem to find the $g(t)$, taken to be real, such that the decoherence is a minimum.

The answer is that, in the limit of large T, i.e., for long pulses, the decoherence is minimal when the pulse shape $g(t)$ is chosen to be

$$g(t) = \sqrt{2}\cos(\pi t/T), \quad \text{for } |t| \leq \tfrac{1}{2}T. \tag{12.33}$$

Furthermore, this minimal decoherence is proportional to $1/T^2$. Note that this pulse shape (12.33) is not only very simple but also independent of the potential $V(x)$.

Since this type of decoherence cannot be avoided even in principle, there is necessarily one more operation on the quantum memory besides write, read, and reset [6]. When a quantum memory is operated on more and more times, the decoherence necessarily accumulates and becomes more and more significant. It is impossible to use scattering to remove or reduce such decoherence.

It is therefore essential to have quantum memories periodically "cleaned up," and this clean-up operation is outside of the operations of writing on, reading, and resetting the memory. In other words, in order for the quantum memory, and hence quantum computing and quantum cryptography, to continue to function, additional operations must be provided, this cleaning-up being one of them.

Consider for definiteness the simplest example where in the memory the two quantum states are degenerate. Suppose an external macroscopic field is applied to this memory to lift the degeneracy; after a time much longer than the lifetime of

the upper state, the memory goes into a pure state. Let the external macroscopic field be removed adiabatically; then the memory remains in a pure state. This is one way in which the clean-up of the memory can be accomplished.

12.6 Quantum cryptography

One of the most interesting and practically important applications of quantum memory is to quantum cryptography, where signals are sent from A (Alice) to B (Bob) with an eavesdropper Eve in between trying to learn what signals are being sent. In the case of quantum key distribution (QKD), as proposed for example in Refs. [14–21], Alice selects a sequence of characters from a finite alphabet and causes a transmitter to send the corresponding signals to Bob. The issue is how much Eve can learn about the signals without letting Alice and Bob know that she is getting that much information.

There are two distinct types of security, conveniently referred to as classical security and quantum security. *Classical security* means that the eavesdropper needs to have much more extensive resources in order to find out what information is sent. The simplest example is that, while it is easy to carry out the multiplication of two large prime numbers, much more computer power is needed to factorize the resulting product into the two prime numbers. By contrast, *quantum security* means that, no matter how much of her resources Eve uses, she *cannot* obtain the desired information without alerting Alice and Bob that there is an eavesdropper.

In the usual proof of quantum security, Eve is assumed to know beforehand all the possible signals that Alice may choose to send. Therefore, if Eve can intercept Alice's transmission and determine with high accuracy which signal it is, then she can retransmit this signal to Bob. In this way, Bob does not know about Eve's intervention, except possibly for a time delay. Such delay is not invoked in the published proofs claiming security of QKD against undetected eavesdropping attacks, i.e., claiming quantum security.

It seems natural that Eve may let the signal from Alice scatter from a quantum memory that has been set beforehand to the standard state s [22]. This is similar to the process of writing discussed in Section 12.2. She can then "read" what is in the quantum memory to ascertain which signal Alice has sent, because in general different signals from Alice lead to different states in the quantum memory. However, this description is oversimplified, because the signal from Alice is usually not an admissible one, meaning that after scattering the state in the quantum memory is no longer a pure state.

At this point, it may be worthwhile to mention the reason why the use of quantum memory requires the reexamination of quantum security in quantum cryptography. The basic reason is that there are more possibilities with the present

formulation. In Section 12.5, the writing, reading, and resetting of a quantum memory have been discussed in some detail. Once a quantum memory has been read, the result can be written on, say, two other quantum memories. This violates the so-called no-cloning theorem [23–25]. This is an example where the use of scattering with one or more spatial variables makes it possible to carry out operations previously considered to be impossible on the basis of models without spatial dimensions.

It should be emphasized that many options are available to Eve, and Eve must choose her action wisely in order to be successful in her endeavor to eavesdrop. Here is the most straightforward use of the quantum memory. This simplest procedure for Eve is to

(1) let the signal from Alice scatter from a quantum memory, which has previously been set to a suitably chosen state;
(2) send in successively N admissible incident waves to scatter from this quantum memory, and record the information from the scattered waves;
(3) find out from this information approximately the state of the memory, which is described by a density matrix, after scattering in step (1);
(4) determine which one of the characters Alice has sent; and
(5) send this character to Bob.

In this way Eve retransmits a quantum state to Bob that, so far as Bob can tell, is what he would have received from Alice without Eve's intervention, except possibly for a time delay. This procedure of Eve is shown schematically in Figure 12.3.

In order to avoid irrelevant complications, the discussion here shall be focused on the QKD two-state protocol B92 of Bennett [16]. In this case, Alice has two signals to choose from, i.e., for each quantum key, Alice can choose either one of the two signals available to her. It is the job of Eve to decide which one of the two Alice has chosen.

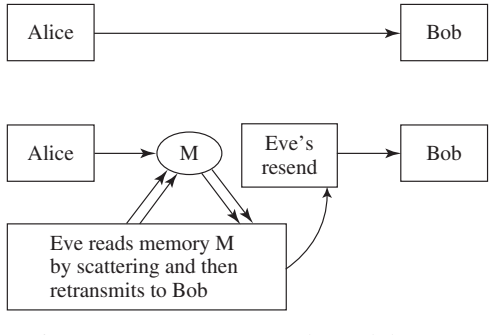


Fig. 12.3 Intercept–resend attack by Eve.

If these two signals available to Alice are orthogonal to each other, then Eve is expected to have no difficulty in determining which one Alice has sent. Therefore, Alice's two signals are always assumed to be non-orthogonal.

The first step is to give a precise description of these two signals. Since Alice and Bob are located at different places (otherwise Bob can merely look over Alice's shoulder), the signals must take the forms of two waves propagating from Alice to Bob. Therefore, the simplest possibility is for the signals to be traveling waves. In the case of one spatial dimension as discussed in Section 12.5, these two signals are described by the real functions

$$\phi_\alpha(t-x), \quad \text{for } \alpha = 1,\ 2. \tag{12.34}$$

It has been assumed that there is no dispersion, and thus the velocity of propagation can be chosen to be 1 without loss of generality.

Let these ϕ_α of (12.34) be normalized so that

$$\int_{-\infty}^{\infty} \mathrm{d}t[\phi_\alpha(t)]^2 = 1, \quad \text{for } \alpha = 1,\ 2. \tag{12.35}$$

Then their inner product is

$$\gamma = \int_{-\infty}^{\infty} \mathrm{d}t\ \phi_1(t)\phi_2(t); \tag{12.36}$$

this γ is the quantity that is chosen to be non-zero.

Instead of attempting to distinguish directly ϕ_1 from ϕ_2 of Alice's possible signals, Eve lets these signals scatter from a quantum memory. Since there is only one spatial dimension as seen from (12.34), it is appropriate to use the model of the quantum memory as described by the Schrödinger equations (12.9). This application of quantum memory to eavesdropping raises two major novel points *vis-à-vis* the case treated so far: (i) the signals used by Alice as described by (12.34)–(12.36) are in general not admissible, and (ii) it is necessary only to distinguish between these two signals. The novel point (i) complicates the task for Eve, while point (ii) makes it easier. Since Alice's signals (12.34) depend on time, the time-dependent Schrödinger equation (12.27) should be used.

In order to use the results of Section 12.3, it is appropriate to apply the Fourier transform to the signals (12.34). Thus define, for $\alpha = 1, 2$,

$$\phi_\alpha(t) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \mathrm{d}k\ f_\alpha(k)\mathrm{e}^{-\mathrm{i}kt}. \tag{12.37}$$

By virtue of (12.35) and (12.36), the $f_\alpha(k)$ have the properties

$$\int_{-\infty}^{\infty} \mathrm{d}k\,|f_\alpha(k)|^2 = 1, \quad \text{for } \alpha = 1,\ 2, \tag{12.38}$$

and

$$\int_{-\infty}^{\infty} \mathrm{d}k\, f_1^*(k) f_2(k) = \gamma. \tag{12.39}$$

Let

$$\begin{bmatrix} a_1^{\mathrm{in}} \\ a_2^{\mathrm{in}} \end{bmatrix}$$

be the initial state in the quantum memory. Then the in field for the scattering of Alice's ϕ_α is, for the even case,

$$\begin{bmatrix} a_1^{\mathrm{in}} \\ a_2^{\mathrm{in}} \end{bmatrix} \phi_\alpha(t+|x|) = (2\pi)^{-1/2} \begin{bmatrix} a_1^{\mathrm{in}} \\ a_2^{\mathrm{in}} \end{bmatrix} \int_{-\infty}^{\infty} \mathrm{d}k\, f_\alpha(k) \mathrm{e}^{-\mathrm{i}k(t+|x|)}. \tag{12.40}$$

It then follows from (12.12) that the out state is given by

$$(2\pi)^{-1/2} \int_{-\infty}^{\infty} \mathrm{d}k\, f_\alpha(k) \mathrm{e}^{-\mathrm{i}k(t-|x|)} S_+(k) \begin{bmatrix} a_1^{\mathrm{in}} \\ a_2^{\mathrm{in}} \end{bmatrix}. \tag{12.41}$$

Unlike the cases treated in Section 12.3, this is *not* a pure state. Averaging over the wave $f_\alpha(k)\mathrm{e}^{-\mathrm{i}k(t-|x|)}$ then shows that the density matrix for the quantum memory after scattering is given by

$$\rho^{\alpha+} = \begin{bmatrix} \rho_{11}^{\alpha+} & \rho_{12}^{\alpha+} \\ \rho_{21}^{\alpha+} & \rho_{22}^{\alpha+} \end{bmatrix} = \int_{-\infty}^{\infty} \mathrm{d}x \text{ [Expression (12.41)][Expression (12.41)]}^\dagger, \tag{12.42}$$

which is explicitly

$$\rho^{\alpha i} = \int_{-\infty}^{\infty} \mathrm{d}k\, |f_\alpha(k)|^2 S_i(k) \begin{bmatrix} a_1^{\mathrm{in}} \\ a_2^{\mathrm{in}} \end{bmatrix} \begin{bmatrix} a_1^{\mathrm{in}*} & a_2^{\mathrm{in}*} \end{bmatrix} S_i^\dagger(k), \tag{12.43}$$

after slight generalization to include the case of the odd wave: $i = +$ or $-$. Note that the right-hand side of (12.43) is independent of t.

The aim of Eve is to distinguish ρ_{1+}^{out} from ρ_{2+}^{out} and/or ρ_{1-}^{out} from ρ_{2-}^{out}.

With a density matrix $\rho^{\alpha i}$, $\alpha = 1$ or 2 and $i = +$ or $-$, in the quantum memory, the next step is to interrogate the memory by scattering in order to determine which of the $\rho^{\alpha i}$ is actually in the memory. In Section 12.4, the required scattering is accomplished using admissible ψ^{in}s; similarly, in general only the admissible ψ^{in}s are to be used here, even though the present case of the density matrix is more complicated.

One might raise the question of why, in the present context, it is desirable to choose the ψ^{in}s to be admissible ones. For the successive scattering process as described by (12.18), the ψ^{in}s of (12.15) have to be admissible so that the content

of the quantum memory remains a pure state. For the present case, although the content of the quantum memory is in general not a pure state after scattering by the signal from Alice, it is nevertheless usually advantageous to use admissible ψ^{in}s. The reason is that using an inadmissible ψ^{in} leads to a loss of information about the content of the quantum memory. It should, however, be kept in mind that there may be circumstances in which this loss of information is not serious and thus non-admissible ψ^{in}s can be used.

For the admissible ψ^{in} to be used to scatter from the quantum memory, the choices consist of a wave number k and also an even or an odd wave. Thus this scattering is described by $S_j(k)$ with $j = +$ or $-$, leading from (12.41) to

$$(2\pi)^{-1/2} \int_{-\infty}^{\infty} \mathrm{d}k'\, f_\alpha(k') \mathrm{e}^{-\mathrm{i}k'(t-|x|)} S_j(k) S_+(k') \begin{bmatrix} a_1^{\text{in}} \\ a_2^{\text{in}} \end{bmatrix}. \tag{12.44}$$

It is the interference between the expressions (12.41) and (12.44) that contains most directly the needed information.

From (12.41)–(12.44), interference leads to the quantity of interest

$$A_\alpha^{ij}(k) = \mathrm{Tr}\,[1 + \mathrm{e}^{\mathrm{i}\theta} S_j(k)]\, \rho^{\alpha i} [1 + \mathrm{e}^{\mathrm{i}\theta} S_j(k)]^\dagger, \tag{12.45}$$

where an extra factor $\mathrm{e}^{\mathrm{i}\theta}$ has been introduced.

From (12.45) it is seen that $A_\alpha^{ij}(k)$ depends not only on the continuous variable k and the $\alpha = 1, 2$ with $i, j = +, -$, but also on the values of a_1^{in}, a_2^{in}, and θ. The task is to choose these available variables such that $A_1^{ij}(k)$ and $A_2^{ij}(k)$ are maximally different so that whether Alice has sent ϕ_1 or ϕ_2 of (12.34) can be efficiently distinguished.

The result of this section challenges the notion that quantum security can follow from inner products alone. In other words, questions are hereby raised about the validity of the usual claim that the γ of (12.36) being non-zero is sufficient to ensure quantum security. While the published analyses have dealt with simplified models of spin-$\frac{1}{2}$ particles and with simplified models of light for a variety of protocols, they have not specifically addressed wavefunctions of particles. Nonetheless, such particles are fair candidates for the analysis of quantum security because such wavefunctions for particles can have any of the inner products that have been featured in previous analyses.

While the lack of quantum security presented here has focused on the QKD two-state protocol B92 of Bennett [16], the reasoning applies to all the known two-, four-, and six-state protocols. The essential assumption is that Eve knows beforehand all the possible signals that Alice may choose to send.

12.7 The Fermi pseudo-potential in one dimension

The above six sections constitute an introduction to the application of the Schrödinger equation to the analysis of some aspects of quantum computing and quantum cryptography, especially quantum memory. Emphasis has been placed on the appearance of spatial variables in the Schrödinger equation, so that the well-developed theory of scattering can be used.

When this present approach was initiated, the Fermi pseudo-potential in one dimension played a central role; it has continued to play an essential role in our understanding of this approach. For this reason, the Fermi pseudo-potential in one dimension will be described briefly in this section.

Since the calculation of the scattering from this case is straightforward but lengthy, there is no point in giving the details here; such details can be found in Ref. [1]. Instead, we shall concentrate on pointing out the salient features and shortcomings.

The Fermi pseudo-potential is well known in three dimensions; it can be written in the form

$$\delta^3(r)\frac{\partial}{\partial r}r, \tag{12.46}$$

as given by Blatt and Weisskopf [26]. The most far-reaching application of this Fermi pseudo-potential is to the study of many-body systems, as initiated by Huang and Yang [27]. For the ground-state energy per particle of a Bose system of hard spheres, the low-density expansion is known to be

$$4\pi a\rho\left[1+\frac{128}{15\sqrt{\pi}}(\rho a^3)^{1/2}+8\left(\frac{4\pi}{3}-\sqrt{3}\right)\rho a^3\ln(\rho a^3)+O(\rho a^3)\right]. \tag{12.47}$$

In this expansion, the second term was first obtained by Lee and Yang [28] using the method of binary collision, but the derivation by Lee, Huang and Yang [29] using the Fermi pseudo-potential is somewhat simpler; the third term, which involves the logarithm, was first obtained by using the Fermi pseudo-potential [30].

As can be seen explicitly from (12.46), the Fermi pseudo-potential is different from zero only at the one point $\mathbf{r}=0$. Indeed, in the mathematics literature, the Fermi pseudo-potential is often called the "point interaction" in three dimensions. This property is taken over also for the case of one dimension [1].

(a) In one dimension, there is the delta-function potential $\delta(x)$ of Dirac, which has this property. Whether this delta-function potential is called a pseudo-potential or not is a matter of terminology. For the present purpose, it is considered to be a Fermi pseudo-potential. Thus the first and simplest pseudo-potential is

$$V_1(x)=g_1\delta(x). \tag{12.48}$$

Let $f(x)$ be a continuous function of x, such as a Schrödinger wavefunction. For the present purposes, instead of $V(x)f(x)$, it is more convenient to consider

$$\int_{-\infty}^{\infty} \mathrm{d}x'\, V(x,x')f(x'). \tag{12.49}$$

For example, if $V(x)$ is a continuous potential, then

$$V(x,x') = V(x)\delta(x-x'). \tag{12.50}$$

For the first pseudo-potential of (12.48), the corresponding $V_1(x,x')$ is

$$V_1(x,x') = g_1\delta(x)\delta(x-x'), \tag{12.51}$$

or equivalently

$$V_1(x,x') = g_1\delta(x)\delta(x'). \tag{12.52}$$

(b) In three dimensions, the potential $\delta^3(\mathbf{r})$ is not acceptable, and this is the reason why the Fermi pseudo-potential takes the form of (12.46). Entirely similarly, in one dimension, the potential $\delta'(x)$ is not acceptable, and instead leads to the second Fermi pseudo-potential. In (12.46), the operator $(\partial/\partial r)r$ serves the purpose of removing a term proportional to $1/r$. Let the second Fermi pseudo-potential in one dimension be written as

$$V_2(x) = g_2\delta'_p(x), \tag{12.53}$$

where what $\delta'_p(x)$ does is

$$\delta'_p(x)g(x) = \delta'(x)\tilde{g}(x), \tag{12.54}$$

where

$$\tilde{g}(x) = \begin{cases} g(x) - \lim\limits_{x\to 0+} g(x), & \text{for } x > 0, \\ g(x) - \lim\limits_{x\to 0-} g(x), & \text{for } x < 0. \end{cases} \tag{12.55}$$

This removes the discontinuity of $g(x)$ at $x = 0$, which is precisely what is needed.

Because of (12.51), the corresponding $V_2(x,x')$ is

$$V_2(x,x') = g_2\delta'_p(x)\delta(x-x'). \tag{12.56}$$

(c) At the beginning of the present investigation it was thought that, in one dimension, there was one pseudo-potential in addition to the Dirac delta-function potential. It was the detailed analysis of the resolvent equation that shows the presence of three independent parameters, and hence three independent Fermi pseudo-potentials including the Dirac delta function.

The expressions (12.50) and (12.51) for the first two pseudo-potentials suggest that the third Fermi pseudo-potential may be given by

$$V_3(x, x') = g_3\delta'_p(x)\delta'_p(x'). \tag{12.57}$$

(This was not the way this third pseudo-potential was originally found; in fact (12.57) was the result of a direct calculation. This calculation also shows that there is no additional pseudo-potential in one dimension.)

It is indeed the appearance of this $V_3(x, x')$ that makes the theory of the one-dimensional Fermi pseudo-potential applicable to the study of quantum memory.

In summary, the three potentials V_1, V_2, and V_3 are given by (12.50), (12.56), and (12.57). Thus the most general Fermi pseudo-potential for the interaction at one point in one dimension is

$$\begin{aligned} V(x, x') &= V_1(x, x') + V_2(x, x') + V_3(x, x') \\ &= g_1\delta(x)\delta(x - x') + g_2\delta'_p(x)\delta(x - x') + g_3\delta'_p(x)\delta'_p(x'). \end{aligned} \tag{12.58}$$

From the experience of working with $\delta'_p(x)$ and the fact that the product of $\delta'(x)$ and a function discontinuous at $x = 0$ is not meaningful, from this point on the convention will be adopted that $\delta'(x)$ always means $\delta'_p(x)$. With this convention, (12.58) is written as

$$V(x, x') = g_1\delta(x)\delta(x - x') + g_2\delta'(x)\delta(x - x') + g_3\delta'(x)\delta'(x'). \tag{12.59}$$

Equation (12.59) can be rewritten in a more symmetric form as follows. Since

$$\delta(x)\delta(x - x') = \delta(x)\delta(x')$$

and

$$\begin{aligned} \delta'(x)\delta(x - x') &= \delta'(x)\delta(x' - x) \\ &= \delta'(x)[\delta(x') - x\delta'(x')] \\ &= \delta'(x)\delta(x') + \delta(x)\delta'(x'), \end{aligned} \tag{12.60}$$

where use has been made of the identity $\delta'(x)x = -\delta(x)$, the general Fermi pseudo-potential (12.59) can be written as

$$V(x, x') = g_1\delta(x)\delta(x') + g_2[\delta'(x)\delta(x') + \delta(x)\delta'(x')] + g_3\delta'(x)\delta'(x'). \tag{12.61}$$

The first and last terms are even while the middle term is odd. That is, under space inversion

$$x \to -x \qquad \text{and} \qquad x' \to -x', \tag{12.62}$$

the coupling constants transform as

$$g_1 \to g_1; \qquad g_2 \to -g_2; \qquad g_3 \to g_3. \tag{12.63}$$

Let this Fermi pseudo-potential (12.61) be used to construct a model for quantum memory. As discussed in Section 12.3, the model is to be based on the two-channel one-dimensional Schrödinger equations (12.9). In view of (12.49), the potential V should be replaced by an integral operator; thus (12.9) take the form

$$\begin{aligned} -\frac{\mathrm{d}^2\psi_1(x)}{\mathrm{d}x^2} + \int_{-\infty}^{\infty} \mathrm{d}x'[V_{11}(x,x')\psi_1(x') + V_{12}(x,x')\psi_2(x')] = k^2\psi_1(x), \\ -\frac{\mathrm{d}^2\psi_2(x)}{\mathrm{d}x^2} + \int_{-\infty}^{\infty} \mathrm{d}x'[V_{21}(x,x')\psi_1(x') + V_{22}(x,x')\psi_2(x')] = k^2\psi_2(x), \end{aligned} \tag{12.64}$$

and (12.8) is replaced by

$$V(x,x') = \begin{bmatrix} V_{11}(x,x') & V_{12}(x,x') \\ V_{21}(x,x') & V_{22}(x,x') \end{bmatrix}.$$

The g_1, g_2, and g_3 in the pseudo-potential (12.61) should be reinterpreted as 2×2 matrices. Furthermore, the condition (12.10) that the potential should be an even function of x implies that g_2 must be zero.

In Section 12.3, the potential must also satisfy the condition that it cannot be diagonalized simultaneously for all x. In the case of the pseudo-potential, this condition translates into the statement that the matrix coefficients of the first and third pseudo-potentials should not commute. Therefore for this application

$$V(x,x') = g_1\delta(x)\delta(x')\alpha_1 + g_3\delta'(x)\delta'(x')\alpha_3, \tag{12.65}$$

where α_1 and α_3 are two 2×2 Hermitian matrices that do not commute. This potential (12.65) has the following nice property: since $g_2 = 0$, the part of the potential $g_1\delta(x)\delta(x')$ does not act on the odd wave, and similarly the part $g_3\delta'(x)\delta'(x')$ does not act on the even wave. The first part of this claim is easy to obtain, and the second part follows from the definition (12.54) of $\delta'_p(x)$.

One of the simplest choices of α_1 and α_3 consists of the Pauli matrices, i.e.,

$$\alpha_1 = \sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \text{and} \qquad \alpha_3 = \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{12.66}$$

This case has been studied in detail in Ref. [1], leading explicitly, from (12.12), to

$$\begin{aligned} S_+(k) &= \frac{(4k^2 - g_1^2) - 4\mathrm{i}g_1k\sigma_1}{4k^2 + g_1^2} = \exp\left[-\mathrm{i}\sigma_1\left(2\tan^{-1}\left(\frac{g_1}{2k}\right)\right)\right] \\ &= \begin{bmatrix} \dfrac{4k^2 - g_1^2}{4k^2 + g_1^2} & \dfrac{-4\mathrm{i}g_1k}{4k^2 + g_1^2} \\ \dfrac{-4\mathrm{i}g_1k}{4k^2 + g_1^2} & \dfrac{4k^2 - g_1^2}{4k^2 + g_1^2} \end{bmatrix} \end{aligned} \tag{12.67}$$

and

$$S_-(k) = \frac{(4 - g_3^2k^2) - 4ig_3k\sigma_3}{4 + g_3^2k^2} = \exp\left[-i\sigma_3\left(2\tan^{-1}\left(\frac{g_3k}{2}\right)\right)\right]$$
$$= \begin{bmatrix} \dfrac{2 - ig_3k}{2 + ig_3k} & 0 \\ 0 & \dfrac{2 + ig_3k}{2 - ig_3k} \end{bmatrix}. \tag{12.68}$$

When neither g_1 nor g_3 is zero, any given element S of SU(2) can be expressed as a finite product of $S_+(k)$ and $S_-(k)$, i.e.,

$$S = S(k_1)\, S(k_2) \cdots S(k_m), \tag{12.69}$$

where each $S(k_i)$ is suitably chosen as $S_+(k_i)$ or $S_-(k_i)$.

What (12.69) means is that, for the model described by (12.66), any unitary transform with determinant equal to 1 can be performed on the quantum memory by a sequence of scattering using the $k_1, k_2, \ldots, k_m$. See also (12.15).

It is natural to ask the following question: how can unitary transforms be carried out if the determinant is not equal to 1? For the choice (12.66), when $\det S \neq 1$, the expression (12.69) cannot hold, because

$$\det S_+(k) = \det S_-(k) = 1, \tag{12.70}$$

as seen from (12.67) and (12.68). Indeed, (12.70) is a consequence of the fact that $\mathrm{Tr}\,\sigma_1 = \mathrm{Tr}\,\sigma_3 = 0$. For more general choices of α_1 and α_3, the first lines of (12.67) and (12.68) still hold, i.e.,

$$S_+(k) = \frac{(4k^2 - g_1^2) - 4ig_1k\alpha_1}{4k^2 + g_1^2} = \exp\left[-i\alpha_1\left(2\tan^{-1}\left(\frac{g_1}{2k}\right)\right)\right],$$
$$S_-(k) = \frac{(4 - g_3^2k^2) - 4ig_3k\alpha_3}{4 + g_3^2k^2} = \exp\left[-i\alpha_3\left(2\tan^{-1}\left(\frac{g_3k}{2}\right)\right)\right]. \tag{12.71}$$

When the traces of α_1 and α_3 are not zero, (12.69) holds for all unitary transforms S.

So far in this chapter, especially in Section 12.3 and the present section, the number of linearly independent quantum states for the memory has been taken to be two. With the present development of the Fermi pseudo-potential in one dimension, there is no longer any difficulty in generalizing to any finite number n of linearly independent quantum states. For example, the coupled one-dimensional Schrödinger equation (12.64) is generalized to be

$$-\frac{d^2\psi_j(x)}{dx^2} + \int_{-\infty}^{\infty} dx' \sum_{j'=1}^{n} V_{jj'}(x, x')\psi_{j'}(x') = k^2\psi_j(x), \quad \text{for } j = 1, 2, \ldots, n, \tag{12.72}$$

and $V(x, x')$ is now an $n \times n$ Hermitian matrix.

The Fermi pseudo-potential (12.65) still holds, except that the matrices α_1 and α_3 are now two $n \times n$ Hermitian matrices that do not commute. When they are properly chosen, and there are many such choices, then the decomposition (12.69) holds also for any n. Indeed, most of the choices are fine, while the unacceptable choices are rather exceptional.

12.8 Conclusion and discussions

The present approach to quantum memory has several new features, and leads to many new directions for further investigation. In this section, we summarize and discuss some of these features, and try to give a partial list of interesting unsolved problems.

(1) Perhaps the most important, novel feature of the present approach to quantum computing and quantum cryptography is the central role played by the Schrödinger equation. In particular, it is realized that spatial variables are of fundamental importance. In the earliest work on quantum computing by Benioff [2] 30 years ago, spatial dimensions were taken into account. However, in much of the later work on this subject, spatial variables have been neglected.

While the Schrödinger equation implies unitary transformations, it contains more information. Its use makes it possible to apply the powerful theory of scattering to perform operations on the quantum memory that are not possible without the spatial dimension. It should perhaps be emphasized that in any case spatial dimensions are physically present.

(2) The analysis of quantum cryptography in Section 12.6 constitutes at most an introduction to this field. It should be noted that Eve has many choices for her purpose of eavesdropping. In particular, there are many different ways Eve can make use of quantum memory. What is described here is by far the simplest way, perhaps it should be described as an oversimplified way. In particular, Eve uses only the density matrix from the content of her quantum memory after scattering by Alice's signal. That γ does not appear emphasizes the fact that the density matrix does not contain all of the information that Eve can retrieve using quantum memory.

(3) In view of this situation, as described under (2), it is tempting to speculate that the following analog from particle physics may prove to be instructive. In particle physics, there are various types of detectors, including in particular scintillators and also chambers such as the time-projection chamber [31]. In the case of a single scintillator, the output when a particle passes through it is typically a single number. For any one of the chambers, the output takes the form of an entire track. The measurement usually considered in the context of quantum computing is analogous to the case of a scintillator in the sense that typically only one number is obtained.

In contrast, the "measurement" as described for example by the scattering sequence (12.15) is closer to that of a chamber, where the output consists of the list (12.19).

In this analogy, the present treatment of quantum cryptography makes use of a chamber only, without any scintillator. The above problem with the absence of the quantity γ in the present approach is perhaps an indication that a better treatment must make use of both scintillators and chambers, as is almost always the case for modern experiments involving high-energy particles. It is likely there will be some interesting developments in the near future in this direction.

(4) The present approach to quantum memory and quantum cryptography by utilizing the Schrödinger equation with at least one spatial variable has many manifestations. Let us mention just one of many that are interesting and important, from both the practical and the theoretical points of view.

As discussed in Section 12.7, with suitable choice of the potential (12.69) holds for all unitary transforms S. Usually, there are many choices for the $S(k_j)$ on the right-hand side, and different choices lead to different values of m. Let $w(S, V; n)$ be the minimum value of m for the given unitary transform S and the potential V, which is an $n \times n$ matrix and may, but need not, be a Fermi pseudo-potential.

From this $w(S, V; n)$, we obtain the following quantities:

$$w(V; n) = \max_S w(S, V; n) \tag{12.73}$$

and

$$W(n) = \min_V w(V; n). \tag{12.74}$$

Thus $w(V; n)$ is the maximum number of terms needed on the right-hand side of (12.69) for any S when V is given, and then the potential V is chosen such that $W(n)$ minimizes this maximum number of terms. It is the asymptotic behavior of this function $W(n)$ for large n that is of importance to quantum computing.

In order to perform a unitary transform S, it is necessary to carry out m scatterings in succession. As can be seen from (12.74), this number of scattering processes can be minimized by a suitable choice of the $n \times n$ potential V. Therefore a unitary transform on the content of a quantum memory involves $W(n)$ steps, not just a single step.

To a large extent, the power of a quantum computer comes from the possibility of operating on a number of quantum memories simultaneously. For example, if there are two linearly independent quantum states in each of N quantum memories, then the number of linearly independent quantum states is 2^N when they are taken together. Therefore, when N quantum memories are operated on together, the number of steps is $W(2^N)$. Since N is a large number in order to make full use of the power of quantum computation, it is the asymptotic behavior of this $W(2^N)$ that determines the speed of quantum computation. If $W(n)$ increases as

a polynomial of n, then the argument of Shor [32] and others remains valid when this number of scattering operations is taken into account. On the other hand, if $W(n)$ increases exponentially with n for large n, then the present method of operating on a quantum memory is likely to cause a significant reduction in the speed of quantum computing.

If $W(n)$ does increase rapidly for large n, then it may be necessary to take the next step in the study of quantum memory. It is at present not known what this step should be, but here is one possibility among many.

In the discussion here of operating on the quantum memory through scattering, the potential $V(x)$ is the same for each of the successive scatterings. For example, in the definition (12.74) of the important function $W(n)$, the minimum is taken over all $V(x)$. There is no reason why this $V(x)$ should be the same for every scattering on the quantum memory. Varying the $V(x)$ seems not only desirable but also natural; however, at present we are unable to offer any idea whatsoever as to how to think about this variation of $V(x)$ from scattering to scattering.

There are perhaps two aspects to this problem. The first one is theoretical: how should the problem of changing $V(x)$ be formulated in a useful way? The second one is practical: what type of variation of $V(x)$ can be performed rapidly? This is just one example of the many future directions of investigation [33].

(5) Finally, it should be emphasized that the present application of quantum mechanics has far-reaching consequences for fundamental aspects of quantum computing and quantum cryptography. These consequences remain to be studied and understood. For example, with writing, reading, and resetting as discussed in Section 12.4, even the way numbers, especially integers, should be efficiently represented in a quantum memory and how arithmetic is to be carried out need to be investigated *ab initio*.

Acknowledgments

For helpful discussions, I am indebted to Howard E. Brandt, Jr., Maurice Jacob, Harold Levine, André Martin, Raymond Stora, Ming Lun Yu, and especially John M. Myers. This work was supported in part by the Air Force Research Laboratory and DARPA under Contract No. F30602-01-C-0170 with BBN Technologies.

References

[1] T. T. Wu and M. L. Yu, Theory and application of Fermi pseudo-potential in one dimension, *J. Math. Phys.* **43**, 5949–5976 (2002).

[2] P. Benioff, The computer as a physical system: a microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines, *J. Statist. Phys.* **22**, 563–591 (1980).

[3] A small sample of papers includes P. Benioff, Quantum mechanical models of discrete processes, *J. Math. Phys.* **22**, 495–507 (1981); D. Z. Albert, On quantum-mechanical automata, *Phys. Lett. A* **98**, 249–252 (1983); R. P. Feynman, Quantum mechanical computers, *Opt. News* **11**, February, 11–20 (1985); D. Deutsch, Quantum theory, the Church–Turing principle and the universal quantum computer, *Proc. Roy. Soc. London A* **400**, 97–117 (1985); S. Parker and M. B. Plenio, Efficient factorization with a single pure qubit and log N mixed qubits, *Phys. Rev. Lett.* **85**, 3049–3052 (2000); and X. Wang, A. Sørensen, and K. Mølmer, Multibit fates for quantum computing, *Phys. Rev. Lett*. **86**, 3907–3910 (2001).
[4] See, for example, R. Pike and P. Sabatier (eds.), *Scattering – Scattering and Inverse Scattering in Pure and Applied Science*, Vols. 1 and 2 (London: Academic Press, 2002).
[5] C. N. Yang and D. Feldman, The S-matrix in the Heisenberg representation, *Phys. Rev.* **79**, 972–978 (1950).
[6] T. T. Wu, Quantum memory: write, read, reset, and decoherence, in *Proceedings of the SPIE, Vol. 5105, Quantum Information and Computation*, E. Donkor, A. R. Pirich, and H. E. Brandt (eds.) (Bellingham, WA: SPIE, 2003), pp. 204–215.
[7] N. N. Khuri and T. T. Wu, New singularities in nonrelativistic coupled channel scattering, I. Second order, *Phys. Rev. D* **56**, 6779–6784 (1997).
[8] N. N. Khuri and T. T. Wu, New singularities in nonrelativistic coupled channel scattering, II. Fourth order, *Phys. Rev. D* **56**, 6785–6797 (1997).
[9] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, Evidence for the 2π decay of the K_2^0 meson, *Phys. Rev. Lett.* **13**, 138–140 (1964).
[10] T. T. Wu and C. N. Yang, Phenomenological analysis of violation of CP invariance in decay of K^0 and $\overline{K}^0$, *Phys. Rev. Lett.* **13**, 380–385 (1964).
[11] BABAR Collaboration, B. Aubert *et al.*, Observation of CP violation in the B^0 meson system, *Phys. Rev. Lett.* **87**, 091801, 1–8 (2001).
[12] Belle Collaboration, K. Abe *et al.*, Observation of large CP violation in the neutral B meson system, *Phys. Rev. Lett.* **87**, 091802, 1–7 (2001).
[13] T. T. Wu and M. L. Yu, Quantum memory: decoherence due to the finite length of the interrogating pulse, Preprint CERN-TH/2002-199, CERN, Geneva (2002).
[14] S. Wiesner, Conjugate coding, *Sigact News* **15**, 78–88 (1983).
[15] C. H. Bennett and G. Brassard, Quantum cryptography: public key distribution and coin tossing, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* (New York: IEEE, 1984), pp. 175–179.
[16] C. H. Bennett, Quantum cryptography using any two non-orthogonal states, *Phys. Rev. Lett.* **68**, 3121–3124 (1992).
[17] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, Generalized privacy amplification, *IEEE Trans. Information Theory* **41**, 1915–1923 (1995).
[18] A. K. Ekert, B. Huttner, G. M. Palma, and A. Peres, Eavesdropping on quantum-cryptographical systems, *Phys. Rev. A* **50**, 1047–1056 (1994).
[19] C. A. Fuchs and A. Peres, Quantum-state disturbance versus information gain: uncertainty relations for quantum information, *Phys. Rev. A* **53**, 2038–2045 (1996).
[20] B. Huttner, A. Muller, J. D. Gautier, H. Zbinden, and N. Gisin, Unambiguous quantum measurement of nonorthogonal states, *Phys. Rev. A* **54**, 3783–3789 (1996).
[21] B. A. Slutsky, R. Rao, P.-C. Sun, and Y. Fainman, Security of quantum cryptography against individual attacks, *Phys. Rev. A* **57**, 2383–2398 (1998).
[22] T. T. Wu, Quantum cryptography and quantum memory, *Proceedings of the SPIE, Vol. 5436, Quantum Information and Computation II*, E. Donkor, A. R. Pirich, and H. E. Brandt (eds.) (Bellingham, WA: SPIE, 2004), pp. 81–92.

[23] D. Dieks, Communication by EPR devices, *Phys. Lett. A* **92**, 271–272 (1982).
[24] W. K. Wootters and W. H. Zurek, A single quantum cannot be cloned, *Nature* **299**, 802–803 (1982).
[25] H. P. Yuen, Amplification of quantum states and noiseless photon amplifiers, *Phys. Lett. A* **113**, 405–407 (1986).
[26] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (New York: J. Wiley, 1952).
[27] K. Huang and C. N. Yang, Quantum mechanical many-body problem with hard-sphere interactions, *Phys. Rev.* **105**, 767–775 (1957).
[28] T. D. Lee and C. N. Yang, Many-body problems in quantum mechanics and quantum statistical mechanics, *Phys. Rev*. **105**, 1119–1120 (1957).
[29] T. D. Lee, K. Huang, and C. N. Yang, Eigenvalues and eigenfunctions of a Bose system of hard spheres and its low temperature properties, *Phys. Rev.* **106**, 1135–1145 (1957).
[30] T. T. Wu, Ground state of a Bose system of hard spheres, *Phys. Rev.* **115**, 1390–1404 (1959).
[31] W. B. Atwood, T. Barczewski, L. A. T. Bauerdick *et al*., Performance of the ALEPH time projection chamber, *Nucl. Instrum. Meth. Phys. Res. A* **306**, 446–458 (1991).
[32] P. W. Shor, Algorithms for quantum computation: discrete logarithms and factoring, *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, (Los Alamitos, CA: IEEE Computer Society, 1994), pp. 124–134.
[33] T. T. Wu, Toward a model for multi-qubit quantum memory, *Proceedings of the SPIE, Vol. 5815, Quantum Information and Computation III*, E. J. Donkor, A. R. Pirich, and H. E. Brandt (eds.) (Bellingham, WA: SPIE, 2005), pp. 62–77.

Index